12-8-2016

3D Subject-Atlas Image Registration for Micro-Computed Tomography Based Characterization of Drug Delivery in the Murine Cochlea

Zhenlin Xu
zx4403@rit.edu

3D Subject-Atlas Image Registration for Micro-Computed Tomography Based Characterization of Drug Delivery in the Murine Cochlea

by

Zhenlin Xu

B.S. Xi'an Jiatong University, 2014

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

December 8, 2016

Signature of the Author _____

Accepted by _____
Coordinator, M.S. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Zhenlin Xu
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science

Dr. Nathan D. Cahill, Thesis Advisor

Dr. David A. Borkholder

Dr. John Kerekes

Date

3D Subject-Atlas Image Registration for Micro-Computed Tomography Based Characterization of Drug Delivery in the Murine Cochlea

by

Zhenlin Xu

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Master of Science Degree
at the Rochester Institute of Technology

Abstract

A wide variety of hearing problems can potentially be treated with local drug delivery systems capable of delivering drugs directly to the cochlea over extended periods of time. Developing and testing such systems requires accurate quantification of drug concentration over time. A variety of techniques have been proposed for both direct and indirect measurement of drug pharmacokinetics; direct techniques are invasive, whereas many indirect techniques are imprecise because they rely on assumptions about the relationship between physiological response and drug concentrations. One indirect technique, however, is capable of quantifying drug pharmacokinetics very precisely: Micro-Computed tomography (μCT) can provide a non-invasive way to measure the concentration of a contrast agent in the cochlea over time. In this thesis, we propose a systematic approach for analyzing μCT images to measure concentrations of the contrast agent ioversol in mouse cochlea. This approach requires segmenting and classifying the intra-cochlea structures from μCT images, which is done via 3D atlas-subject registration to a published atlas of the mouse cochlea. Labels of each intra-cochlear structure in the atlas are propagated through the registration transformation to the corresponding structures in the μCT images. Pixel intensities are extracted from three key intra-cochlea structures: scala tympani (ST), scala vestibuli (SV), scala media (SM) in the μCT images, and these intensities are mapped into concentrations using a linear model between solution concentration and image intensity

that is determined in a previous calibration step. To localize this analysis, the ST, SV, SM are divided into several discrete components, and the concentrations are estimated in each component using a weighted average with weights determined by solving a nonhomogeneous Poisson equation with Dirichlet boundary conditions on the component boundaries. We illustrate this entire system on a series of μCT images of an anesthetized mouse that include a baseline scan (with no contrast agent) and a series of scans after injection of ioversol into the cochlea.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Cahill for the continuous support of my study and related research, for his patience, encouragement, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my M.S study.

I also would like to thank Dr. Borkholder whose project my research is applied for, for his insightful comments and motivation, also for showing me more real meaning of my researching from perspective of a different area.

My sincere thanks also goes to Dr. Kerekes, who served on my thesis committee, and Dr. Linte, who offered me space in his lab, and all faculty and staff that helped me in my research at RIT. Without their support it would not be possible to conduct this research.

I thank my fellow classmates and research-mates at CIS and other programs for the stimulating discussions, for the time we were working together before deadlines, and for all the fun we have had in the last two years.

Last but not the least, I would like to thank my parents for supporting me spiritually throughout writing this thesis and throughout my life in general.

To My Parents

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations and Related Work

Local delivery of medicine treatment to the inner ear is a potential way to treat hearing impairment or loss. However, it is challenging to developing drug delivery systems and therapeutics due to the difficulty of obtaining direct measurements of drug concentration in the inner ear over time. Some techniques have used computer simulation to predict drug pharmacokinetics without having to rely on direct measurements. 1D [35] and 3D [27] simulation models of cochlea have been built and used to interpret published data on the time course of gentamicin concentration in chinchilla cochlea[28] and corticosteroid concentration in guinea pig cochlea [26].

Various approaches have been proposed to quantify drug pharmacokinetics in the cochlea. Direct measurement of drug concentration, either using ion-selective microelectrodes [36] or sampling from the basal or apical turn of the cochlea[1], are invasive and will influence the concentration of drugs. Indirect methods provide estimates of drug concentration based on the functional assessment of hearing [6, 23]. These methods assume that the physiological response of the injected drug is proportional to the drug concentration in the cochlea and that cochlear hair cell sensitivity is uniform. The spatial resolution of concentration measurement relies on the density of electrodes and the sampling posi-

tions for direct measurement, while the indirect methods estimate the concentrations as a whole without spatial distribution. Kanzaki *et al.* established an in vivo imaging system to monitor drug delivery utilizing luciferase-luciferin reaction in the cochlea of GFAP-Luc transgenic mice [18]. Luciferin delivered into the inner ears reacted with the luciferase in GFAP-expressing cells in the cochlear spiral ganglion generating photon bioluminescence and the length of bioluminescence was imaged to estimate the drug concentrations. With this approach, it was found that the volume of a drug, not its concentration, is the important factor for inner ear drug delivery when the drugs are injected systemically [19]. However this approach cannot measure the drug concentration quantitatively.

Because of the limitations of existing approaches, micro-computed tomography (μCT) imaging, a non-invasive method, has been explored for visualizing cochlear structures and tracing the kinetics of drugs. For example, 3D modeling of the human cochlea was done based on μCT data [8, 31]. Zou *et al.* used μCT with $21.9\mu m$ resolution to visualize the silver nanoparticles and measure their concentrations in the cochlea [45].

One of the critical problems in quantifying drug pharmacokinetics via imaging is segmenting and identifying intra-cochlear structures from images of the cochlea. It is especially difficult to distinguish the membranes between different fluid-filled scala. Some studies applied advanced algorithms like region growing [43], active contours [44], and level sets [42], but their segmentation results can not differentiate between individual scalae. A recent study based on active shape models [24] used a high resolution statistical model trained from several μCT image sets to segment the new CT image. However, this approach requires model training from manually segmented cochlea image sets. A recent study by Haghpanahi *et al.* [14] used atlas-based segmentation to localize the intra-cochlea structures by constructing an atlas, performing 2D registration, propagating segmentation labels to μCT. However, 2D registration is still limited in accuracy, making it difficult to measure drug pharmacokinetics at high spatial resolution.

1.2 Objectives

We propose to build off of the technique of Haghpanahi *et al.* [14] by extending the atlas-based segmentation to a fully 3-D process, and by robustly extracting pharmacokinentics curves from 3D segmented scalae. As a result, the objectives of this thesis are:

- Propose a systematic approach for analyzing 3D μCT images to measure the concentrations of the contrast agent ioversol in mouse cochlea.
- Develop and implement an accurate and robust 3D subject-atlas registration algorithm using the Insight Segmentation and Registration Toolkit (ITK) to segment and classify the intra-structures of cochlea in 3D μCT images.
- Extract and quantify concentrations in scalae tympani (ST), scala vestibuli (SV) and scala media (SM) from 3D segmented μCT images of cochlea.

1.3 Thesis Overview

The following chapters of this thesis are organized as follows:

Chapter 2 gives a background on atlas-based segmentation and a review of key components of the regular intensity-based image registration framework. It also introduces a method for constructing shape representations by solving the Poisson Equation.

Chapter 3 describes the methodology of our whole system including μCT imaging, atlas-subject registration, to concentration extraction.

Chapter 4 illustrates our system using successful segmentation and concentration quantification results.

Chapter 5 summarizes the thesis with a discussion of future work.

Chapter 2

Background

This chapter will introduce the key technique in this thesis: atlas-based segmentation, which turns the segmentation task into a registration problem. Then it will describe a regular intensity-based registration workflow and its components, including similarity metric, optimizers and transformations. In addition, it describes a shape representation approach by solving a nonhomogeneous Poisson equation that will be used to determine weights for quantifying average concentrations of contrast agent in different structures of the cochlea.

2.1 Atlas-based Image Segmentation and Image Registration

Segmentation of medical images is a challenging problem because of many issues, including the low contrast of images, fuzzy object contours and similar intensities with adjacent objects of interests [2]. Using a reference image, called an atlas, with prior knowledge of the subject in the image to be segmented could help in the segmentation task. The prior knowledge is commonly represented as label maps.

With a labeled atlas, the segmentation problem is reduced to an image registration problem by propagating the transformation found by registration process to atlas labels. To segment a new image, the transformation that maps the atlas to the subject image

needs to be computed by registration. This transformation is then applied to the labels assigned to structures from the atlas onto the subject image to be segmented.

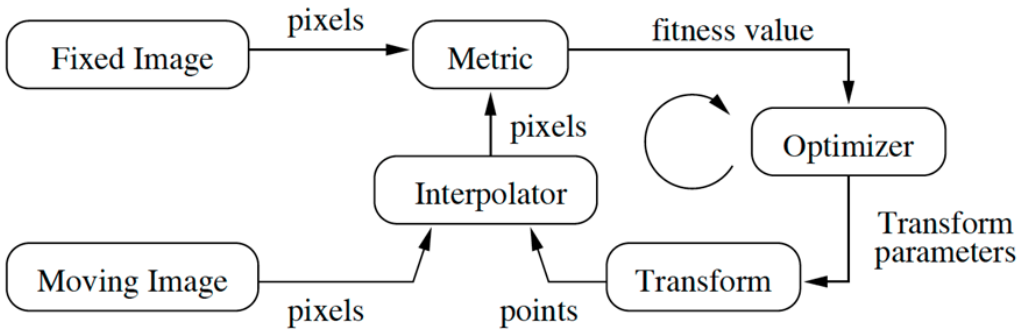


Figure 2.1: The typical intensity-based registration workflow [17]. In our case, a high resolution atlas works as the moving image and the lower resolution micro-CT image works as the fixed image.

Image registration is the process of determining the spatial transform that maps points from one image to homologous points in the second image [17]. Our registration algorithm is based on the typical intensity-based registration workflow shown in Fig. 2.1 [17]. There are two images that are the input data to the registration process. One is referred to the *fixed* or *reference* image $f(\mathbf{X})$ and the other is referred to as the *moving* or *target* image $m(\mathbf{X})$, where $\mathbf{X}$ is the pixel position in N-dimensional space. Registration aims to find the best mapping that aligns the moving image to the fixed image, and the transformation T represents the mapping from the fixed image space to the moving image space. A hierarchy of transformations will be introduced in section 2.4. The *interpolator* is used to evaluate moving image intensities at non-grid positions. The similarity *metric* $M(f, m, T)$ measures how well the transformed *moving* image is aligned with the *fixed* image. Section 2.2 presents

several popular similarity metrics. If we define the *metric* so that lower values mean greater similarity between the transformed moving image and the fixed image, registration can be treated as a optimization problem that solves:

$$T^* = \arg\min M(f, m, T), \tag{2.1}$$

where the T^* is the optimal transformation. Section 2.3 includes more details about optimizers for image registration.

2.2 Similarity Metric for Image Registration

In the registration framework shown in Fig. 2.1, the similarity metric is perhaps the most critical component. Similarity metrics quantify how well the transformed moving image fits the fixed image from the pixel intensity of the images, according to which the optimal transformation T is iteratively determined by the optimizer. However, the selection of metric to use highly depends on the registration problem itself, and there are no clear-cut rules. An important factor when selecting similarity metrics is the set of imaging modalities involved in the registration. Some metrics, for example, the Sum of Squared Differences, are only suitable for comparing images obtained from the same imaging modality, while others, like Mutual Information, can handle inter-modality comparisons. Several popular similarity metrics are introduced in this subsection.

2.2.1 Sum of Squared Differences

One of the simplest similarity metrics is the *Sum of Squared Differences* (SSD) . Between image A and B within an overlap region,

$$SSD = \frac{1}{N} \sum_{x_i \in \Omega_{A,B}} (A(x_i) - B(x_i))^2, \tag{2.2}$$

where x_i is the ith pixel location in the overlap region $\Omega_{A,B}$ comprising N pixels.

The value of this SSD metric is zero only when two identical image are registered

perfectly. Misalignment between images A and B result in large values. This metric relies on the assumption that intensities representing the same homologous point must be the same in both images. Thus the use of SSD is restricted to inter-modality registration. Moreover, any linear change in the intensity results in a high SSD value. The SSD metric is very sensitive to a small number of pixels that have very large intensity differences between images A and B. This might arise, for example, if a contrast agent is injected into the patient between the imaging of A and B, which happens in our experiments. Therefore, SSD is not a good choice for our problem.

2.2.2 Normalized Correlation Coefficient

Normalized Correlation Coefficient (NCC) is a better similarity metric for handling registration in which there is a linear relationship between the intensity values in the fixed and moving images:

$$NCC = \frac{\sum_{x_i \in \Omega_{A,B}} (A(x_i) - \bar{A}) \cdot (B(x_i) - \bar{B})}{\left\{ \sum_{x_i \in \Omega_{A,B}} (A(x_i) - \bar{A})^2 \cdot \sum_{x_i \in \Omega_{A,B}} (B(x_i) - \bar{B})^2 \right\}^{1/2}}, \tag{2.3}$$

where $\bar{A}$ and $\bar{B}$ are the mean pixel values within the region $\Omega_{A,B}$ in images A and B, respectively. The optimal value of this metric is one, and poor matches between the images results in small values. This metric is also limited to be used in registration of images from the same imaging modality. Compared to SSD, NCC is insensitive to multiplicative factors between the two images. In our application, the μCT image intensities are in Hounsfield units (HU), stored in 16bits integers ranging from -1000 to tens of thousands, while the atlas is an 8-bits gray-scale image derived from a different μCT scan. Therefore, NCC is a suitable similarity metric for our case.

2.2.3 Mutual Information

Mutual Information (MI) is a popular similarity metric for multi-modality registration that emerges from information theory. The definition of mutual information is based on

entropy, which is a measure of the information content or complexity of a random variable. The entropy of image A is defined by:

$$H(A) = -\sum_{a} p_A(a) log p_A(a), \tag{2.4}$$

where a represents intensity bins of image A and $p_A(a)$ is the probability mass function (p.m.f.). Similarly, the entropy of image B is:

$$H(B) = -\sum p_B(b) log p_B(b). \tag{2.5}$$

Joint entropy measures the information of combined images A and B:

$$H(A,B) = -\sum_{a}\sum_{b} p_{AB}(a,b) log p_{AB}(a,b), \tag{2.6}$$

where $p_{AB}(a,b)$ is the joint p.m.f. of images A and B. If A and B are independent, then

$$p_{AB}(a,b) = p_A(a) p_B(b), \tag{2.7}$$

and the joint entropy will be the sum of entropies of individual images:

$$H(A,B) = H(A) + H(B). \tag{2.8}$$

If any dependency exists between images A and B, then

$$H(A,B) < H(A) + H(B), \tag{2.9}$$

and the more similar A and B are, the lower their joint entropy. The difference between the sum of the individual entropy and the joint entropies is called Mutual information:

$$I(A,B) = H(A) + H(B) - H(A,B). \tag{2.10}$$

Therefore, greater similarity between images results in lower joint entropy and thus higher mutual information. The terms $H(A)$ and $H(B)$, called marginal entropies, are penalties that bring lower value to $I(A, B)$ when the overlap between A and B contains only background, which causes failed registration. To address this problem, Normalized Mutual Information (NMI) was proposed [41]:

$$NMI(A, B) = 1 + \frac{I(A, B)}{H(A, B)} = \frac{H(A) + H(B)}{H(A, B)}, \tag{2.11}$$

which is less sensitive to changes of overlap. Mutual information and Normalized Mutual Information measure the dependence between two images and does not rely on the one-to-one intensity correspondence. Therefore, they are well suited for multi-modality registration[29].

2.3 Optimization in Image Registration

As mentioned before, the registration framework in Fig. 2.1 is solving an optimization problem that determines the transformation that optimizes the similarity metric. Optimization problems are typically solved iteratively. If we parameterize the transformation T with the parameter vector $\mathbf{p}$ as $T(\mathbf{p})$, then the parameter space of a rigid transformation in 3D has six dimensions, whereas the parameter space for an affine transformations have twelve dimensions. Deformable registration has more, often many hundreds or thousands, degrees of freedom, and correspondingly the parameter space has a very high number of dimensions. Each point in the parameter space corresponds to a distinct transformation.

The job of the optimizer is finding the optimal location in the parameter space. However, searching a parameter space for image registration is usually not simple. There are often multiple local optima within the parameter space, and registration can fail if the optimization algorithm converges to the wrong local optimum. Some of these local optima may be very small, caused either by interpolation artifacts, or a local good match between features or intensities. The "local optimum" problem challenges most optimization algorithms. Approaches to mitigate this problem involve blurring the images prior to

registration or using the multi-resolution strategy introduced in section 3.3.1.

To determine the optimal parameter vector $\mathbf{p}$, the optimizer starts with an initial estimate of the transformation which is gradually refined iteratively. The optimizer calculates the value of similarity metric using the current estimate of the transformation; then, it updates the estimate of the transformation and evaluates the similarity metric value again. The loop continues until the algorithm converges, when no transformation that results in a better value of the similarity metric can be found, within a preset tolerance. This iterative optimization strategy could be described as:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \mathbf{d}_k, \tag{2.12}$$

in which $\mathbf{p}_k$ is the parameter vector at the kth iteration, $\mathbf{d}_k$ is the search direction at iteration k, and a_k is a scalar factor controlling the step size along the direction of $\mathbf{d}_k$. The optimization methods varies in the way of computing $\mathbf{d}_k$ and a_k.

In the following sections, several optimization approaches are introduced and compared in three categories. The first category is deterministic gradient-based algorithms, including gradient descent [9], quasi-Newton [14], nonlinear conjugate gradient [15]. These algorithms determine the search direction $\mathbf{d}_k$ based on the gradient of the cost function $\mathbf{g}(\mathbf{p}) = \nabla M(\mathbf{p})$ and assume that $\mathbf{g}(\mathbf{p})$ can be computed exactly. The second category is stochastic gradient-based algorithms. Instead of computing exact derivative, only stochastic approximations of the gradient are used to make the algorithm faster. The third category is evolutionary algorithms, which determine the searching direction based on a random process.

2.3.1 Deterministic Gradient-based Optimization

The *gradient descent* method takes steps in the negative gradient direction:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - a_k \mathbf{g}_k, \tag{2.13}$$

where $\mathbf{g}_k$ is the gradient of the similarity metric evaluated at the current iteration k.

Assuming the sequence $\{\mathbf{p}_k\}$ converges to a local optimal solution $\hat{\mathbf{p}}$, it has been established in [4] that

$$\exists K \geq 0 \text{ and } \rho > 0, \quad \frac{\|\mathbf{p}_{k+1} - \hat{\mathbf{p}}\|}{\|\mathbf{p}_k - \hat{\mathbf{p}}\|} \leq \rho, \quad \text{for all } k \geq K, \tag{2.14}$$

which means that gradient descent exhibits a linear convergence rate.

There are variants of the gradient descent method in the way of determining the learning rate a_k, for example, using a decaying function: $a_k = a/(k+A)^\alpha$ where $a > 0$, $A \geq 1$, and $0 \leq \alpha \geq 1$.

Our algorithm uses a variant of gradient descent that attempts to avoid taking steps that are too large: Regular Step Gradient Descent Optimizer [17]. When the direction of $\mathbf{g}_k$ changes abruptly, the optimizer assumes that a local minimum has been passed and reduces the step size by a relaxing factor which is preset by user.

Quasi-Newton (QN) methods [25] are inspired by Newton's method that uses the gradient and the Hessian matrix of the cost function:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - H_k^{-1}\mathbf{g}_k, \tag{2.15}$$

where H_k is the Hessian matrix of cost function at the current kth iteration. Since the Hessian matrix contains the second-order information, Newton's method can converge extremely quickly. However, the cost of computing the Hessian matrix and its inverse is large especially when $\mathbf{p}$ is of high dimensionality, for example in the deformable registration problem. QN methods solves this problem by using an approximation to the inverse Hessian matrix: $B_k \approx H_k^{-1}$ instead of exact solution. A line search algorithm is commonly used to determine the step size a_k. QN methods are given by :

$$\mathbf{p}_{k+1} = \mathbf{p}_k - a_k B_k \mathbf{g}_k. \tag{2.16}$$

Given certain conditions, many QN methods can be shown to be superlinearly convergent

[10]:

$$\lim_{k\to\infty} \frac{\|\mathbf{p}_{k+1} - \hat{\mathbf{p}}\|}{\|\mathbf{p}_k - \hat{\mathbf{p}}\|} \to 0. \tag{2.17}$$

Many approaches were proposed to update B_k, including Davidon-Fletcher-Powell (DFP), and Broyden-Fletcher-Goldfarb- Shanno (BFGS), Symmetric-Rank-1 (SR1) and Broyden family[25]. The BFGS method is the most favored one and its rule of updating B_k is:

$$B_{k+1} = (I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k})H_k(I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}) + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}, \tag{2.18}$$

where I is the identity matrix, $\mathbf{s}_k = \mathbf{p}_{k+1} - \mathbf{p}_k$ and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$.

An inexact line search algorithm is used in QN methods to compute the step size a_k. A set of inequalities called *the strong Wolfe conditions* is used in the line search procedure:

$$M(\mathbf{p}_{k+1}) \leq M(\mathbf{p}_{k+1}) + c_1 a_k \mathbf{d}_k^T \mathbf{g}_k, \tag{2.19}$$

$$\left|\mathbf{d}_k^T \mathbf{g}_{k+1}\right| \leq \left|c_2 \mathbf{d}_k^T \mathbf{g}_k\right|, \tag{2.20}$$

with $0 < c_1 < c_2 < 1$, where d_k is the search direction of the optimizer. The first inequality (2.19) is called the sufficient decrease or Armijo condition, which ensures that the step size a_k decreases the cost function M sufficiently. The second inequality 2.20 is called the curvature condition, which enforces reasonable progress towards a stationary point of the cost function.

Conjugate gradient methods are among the most useful techniques for solving large linear system of equations while they can be adapted to solve nonlinear optimization problems. The nonlinear conjugate gradient method also follows the iterative scheme in (2.12). Similar to QN methods, the step size a_k can be computed by line search algorithms. The search direction $\mathbf{d}_k$ is updated differently, however, by:

$$\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}. \tag{2.21}$$

The four most well-known methods for computing β_n are:

$$\text{Fletcher-Reeves method:[11]} \quad \beta_k^{FR} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}, \tag{2.22}$$

$$\text{Polak-Ribière [30]} \quad \beta_k = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}, \tag{2.23}$$

$$\text{Hestenes-Stiefel method [16]} \quad \beta_k = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{d}_{k-1}^T (\mathbf{g}_k - \mathbf{g}_{k-1})}, \tag{2.24}$$

$$\text{Dai-Yuan method [9]} \quad \beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{d}_{k-1}^T (\mathbf{g}_k - \mathbf{g}_{k-1})}, \tag{2.25}$$

The convergence properties depend on the line search algorithms; it is shown [22] that a superlinear rate of convergence can be achieved with a more practical inexact line search routine.

2.3.2 Stochastic Gradient-based Optimization

The stochastic gradient descent (SGD) method also minimizes the cost function via an iterative procedure, in a similar manner to the deterministic gradient method. But instead of computing the exact gradient $\mathbf{g}_k$ at each iteration, SGD methods use an approximation of $\tilde{\mathbf{g}}_k$ for the gradient and define a decaying sequence $\{a_k\}$:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - a_k \tilde{\mathbf{g}}_k. \tag{2.26}$$

A well known SGD method was proposed by Robbin and Monro [32]. It assumes that an approximation of the derivative of the cost function is available:

$$\tilde{\mathbf{g}}_k = \mathbf{g}_k + \epsilon_k. \tag{2.27}$$

It guarantees convergence to the solution $\hat{\mathbf{p}}$ if the bias of the approximation error ϵ_k goes to zero:

$$\lim_{k \to \infty} E(\tilde{\mathbf{g}}_k) = \mathbf{g}_k, \tag{2.28}$$

where $E(.)$ donates expectation. Note that, if $\epsilon_k = 0$ in every iteration, the SGD method is equivalent to the DGD method.

The gradient approximation $\tilde{\mathbf{g}}_k$ does not neccessarily vanish closely to the solution $\hat{\mathbf{p}}$ where $\mathbf{g}(\hat{\mathbf{p}}) = 0$. Therefore, the decaying sequence must satisfy that $\lim_{k\to\infty} a_k = 0$ to guarantee the convergence of $\mathbf{p}_k$. Most commonly, a_k is defined as

$$a_k = a/(k + A)^{\alpha} \tag{2.29}$$

where $a > 0$, $A \geq 1$ and $0 \leq \alpha \geq 1$ are user defined constants.

A stochastic gradient descent method is often applied when computation of the exact gradient is very costly. Using an approximation of the exact gradient could decrease the computation time per iteration but may have negative effects on the speed of convergence [20].

2.3.3 Evolutionary algorithm

Evolution Strategies (ESs) are algorithms that imitate the principles of natural evolution in order to solve the optimization problem [3]. These algorithms can be expressed in the language of biology as follows:

Step 1: (Initialization)
A given population consists of μ individuals. Each is individually characterized by its genotype consisting of n genes, which unambiguously determine the fitness for survival.

Step 2: (Mutation) By mutation and recombination operations, each individual parent produces λ/μ offspring on average, so that a total number of λ offspring individuals are available.

Step 3: (Selection) Select the μ best of the λ offspring to form parents of the following generation and continue at Step 2.

When $\mu = \lambda = 1$, the simplest imitation of evolution is called (1+1) evolutionary strategy. For our optimization problem (2.1), the algorithm could be formalized as follows:

Step 1: (Initialization) Define an initial position in the parameter space $\mathbf{p}_0$ that represents an initial transformation T_0.

Step 2: (Mutation) An offspring of the current parameter vector is generated by

$$\mathbf{p}_k^o = \mathbf{p}_k + a_k \mathbf{d}_k, \tag{2.30}$$

where the search direction $\mathbf{d}_k$ is generated from a normal distribution:

$$\mathbf{d}_k \sim \mathcal{N}(0, I). \tag{2.31}$$

Step 3: (Selection) The next generation of the parameter vector is determined by

$$\mathbf{p}_{k+1} = \begin{cases} \mathbf{p}_k^o, & \text{if } M(\mathbf{p}_k^o) < M(\mathbf{p}_k) \\ \mathbf{p}_k, & \text{otherwise} \end{cases}, \tag{2.32}$$

and the step size is updated by

$$a_{k+1} = \begin{cases} a_k \cdot c_{increase}, & \text{if } M(\mathbf{p}_k^o) < M(\mathbf{p}_k) \\ a_k \cdot c_{decrease}, & \text{otherwise} \end{cases}, \tag{2.33}$$

where $c_{increase}$ and $c_{decrease}$ are used defined scalar. Continue to step 2 until convergence condition is satisfied.

For evolutionary strategies with more than one parent and one generation and different variants, [3] and [38] provide good reviews.

2.4 Hierarchy of Transformation of Registration

The type of deformation between the moving image and the fixed image determines the complexity of the registration problem and also the choice of registration methods. In

this section, we describe a hierarchy of transformations and their algebraic representations. These transformations usually differ in degrees of freedom, with more complex transformations having higher degrees of freedom. The typical transformations used in image registration are: rigid, similarity, affine, projective and deformable. Their 2D and 3DãĂĂexamples are shown in Fig. 2.2 and 2.3 along with their corresponding degrees of freedom. Since the projective transformation is more popular in registration of RGB images captured by projective cameras but rarely appears in medical image registration, we only introduce the other four types of transformations in the following sections. Every type of transformation will be introduced in both $\mathbb{R}^2$ and $\mathbb{R}^3$.

2.4.1 Rigid Transformation

Rigid transformations, or isometries, are transformations that preserve Euclidean distance. In $\mathbb{R}^2$, a rigid transformation is formulated as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \epsilon\cos\theta & -\sin\theta & t_x \\ \epsilon\sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{2.34}$$

where $\epsilon = \pm 1$. If $\epsilon = 1$, then the isometry is orientation preserving and is a Euclidean transformation (a composition of a rotation and translation). If $\epsilon = -1$ then the isometry reverse the orientation, for example the composition of a reflection and an Euclidean transformation. Note that some references define rigid transformations solely with the choice $\epsilon = 1$. We follow the convention of [15] and allow rigid transformations to include the case where $\epsilon = -1$ as well.

A 2D Euclidean transformation can also be written more concisely as:

$$\mathbf{x}' = T_{\text{rigid}}(\mathbf{x}) = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0^T} & 1 \end{bmatrix} \mathbf{x}, \tag{2.35}$$

where $\mathbf{x} = (x, y, 1)$, $\mathbf{x}' = (x', y', 1)$, $\mathbf{R}$ is a 2×2 rotation matrix, $\mathbf{t}$ a translation 2-vector, and $\mathbf{0}$ a null 2-vector. A Euclidean transformation in $\mathbb{R}^3$ can be also described in the form

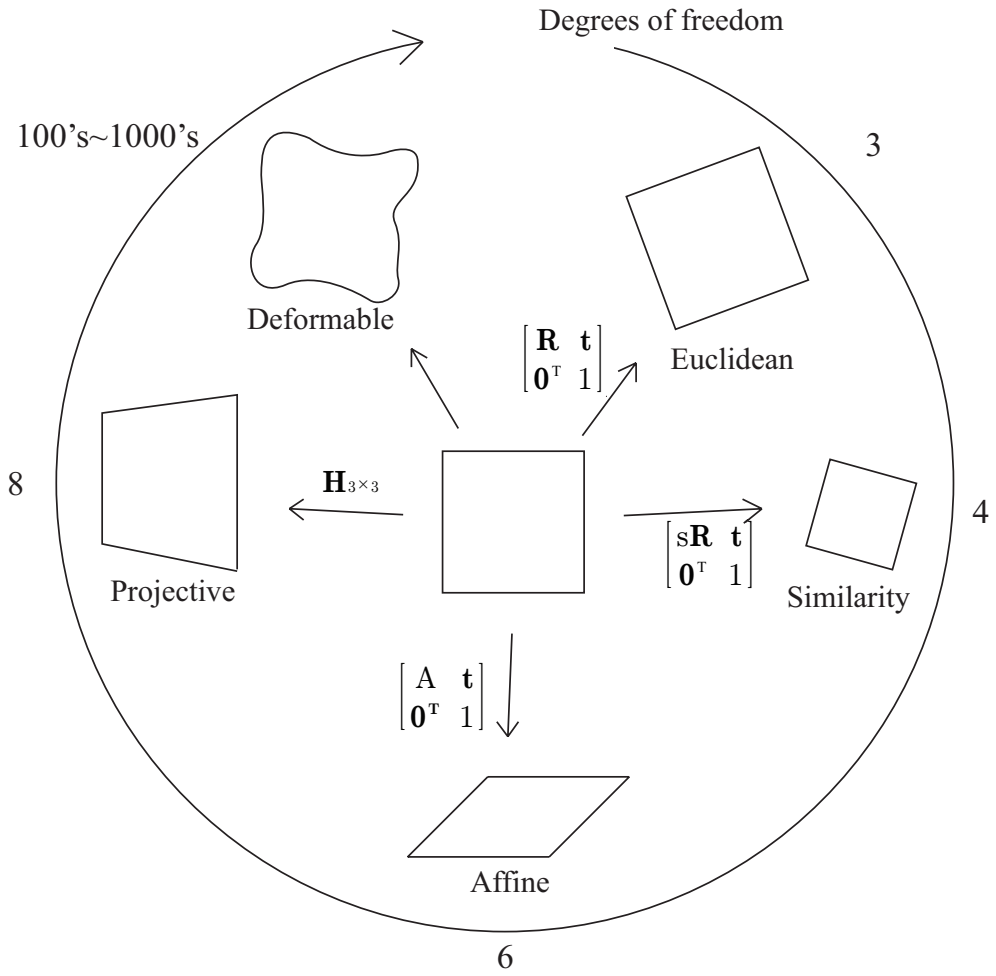


Figure 2.2: Hierarchy of Geometrical Transformations in 2D

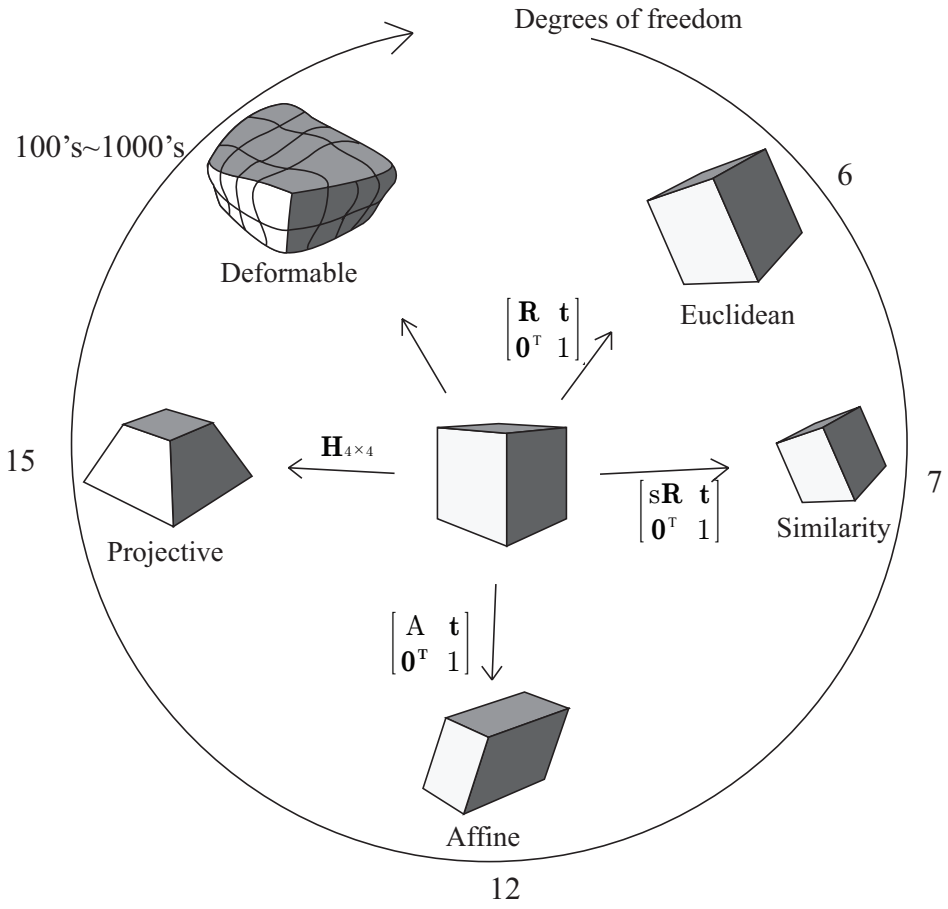


Figure 2.3: Hierarchy of Geometrical Transformations in 3D

of (2.3) where $\mathbf{x} = (x, y, z, 1)$, $\mathbf{x}' = (x', y', z', 1)$, $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ a translation 3-vector, and $\mathbf{0}$ a null 3-vector.

A 2D Euclidean transformation has three degrees of freedom: one for the rotation and two for the translation. Thus, at least three parameters must be specified in order to define the transformation. A 3D Euclidean transformation has six degrees of freedom, three for the rotation and three for the translation, and at least six parameters are necessary to define the transformation.

2.4.2 Similarity Transformation

A similarity transformation can be seen as a rigid transformation combined with an isotropic scaling. In the case of a 2D Euclidean transformation composed with a scaling (i.e. no reflection), a 2D similarity transformation can be represented as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{2.36}$$

and a 3D similarity transformation can be written in block form as:

$$\mathbf{x}' = T_{\text{similarity}}(\mathbf{x}) = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0^T} & 1 \end{bmatrix} \mathbf{x}, \tag{2.37}$$

where $\mathbf{x} = (x, y, z, 1)$, $\mathbf{x}' = (x', y', z', 1)$, s is the isotropic scaling factor, $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ a translation 3-vector, and $\mathbf{0}$ a null 3-vector. Because of the scaling factor, a similarity transformation have one more degree of freedom than a Euclidean transformation that is four degrees of freedom in $\mathbb{R}^2$ and seven degrees of freedom in $\mathbb{R}^3$. A similarity transformation still preserves the ratio of lengths and angles.

2.4.3 Affine Transformation

An affine transformation is a non-singular linear transformation followed by a translation [15]. It adds scaling and shearing to a rigid transformation. An affine transformation in $\mathbb{R}^2$ is represented as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{2.38}$$

and in $\mathbb{R}^3$ as:

$$\mathbf{x}' = T_{\text{affine}}(\mathbf{x}) = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0^T} & 1 \end{bmatrix} \mathbf{x}, \tag{2.39}$$

where $\mathbf{A}$ is a 3×3 non-singular matrix. The degrees of freedom of an affine transformation are six in 2D and twelve in 3D. An affine transformation preserves the parallelism of lines but not angles or lengths.

2.4.4 Deformable Transformation

The above transformations are often regarded as a global transformation and preserve the straightness of lines. To model more complex deformations when images show high levels of anatomical variability between subjects, for example, the local variant between the cochlea atlas and actual imaged cochlea, more detailed and localised deformable transformations are essential. The literature [40] provides a good view of local deformation models.

A deformable transformation can be described as:

$$\mathbf{x}' = T_{deformable}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x}), \tag{2.40}$$

where $\mathbf{x} = (x, y, z)$ and $\mathbf{x}' = (x', y', z')$ are the coordinate vectors of points before and after transformation, respectively, and $\mathbf{u}(\mathbf{x})$ is the displacement field.

Free-Form Deformations (FFDs) are among the most common types of deformation models in medical image registration. They were first developed in the computer graphics community [39] and then widely used in the medical image analysis community [21, 34].

FFDs deform an object by manipulating an underlying mesh of control points, producing a smooth transformation. This requires a regular mesh of control points with uniform spacing. Let the rectangular mesh grid in a image with size $N_x \times N_y \times N_z$ denoted by Φ with $K_x \times K_y \times K_z$ control points ($K_x \ll N_x, K_y \ll N_y, K_z \ll N_z$) with uniform spacing δ. The deformation field is given as a summation of tensor products of spline functions, a popular choice of which is the cubic B-spline:

$$\mathbf{u}(\mathbf{x}) = \sum_{l}^{3}\sum_{m}^{3}\sum_{n}^{3} B_l(u)B_m(v)B_n(w)\mathbf{d}_{i+l,j+m,k+n}, \tag{2.41}$$

where $\mathbf{d}_{i,j,k}$ is a control point of grid Φ, B_l represents the lth cubic B-spline basis functions:

$$B_0(u) = \frac{(1-u)^3}{6}, \tag{2.42}$$

$$B_1(u) = \frac{(3u^3 - 6u^2 + 4)}{6}, \tag{2.43}$$

$$B_2(u) = \frac{(-3u^3 + 3u^2 + 3u + 1)}{6}, \tag{2.44}$$

$$B_3(u) = \frac{u^3}{6}, \tag{2.45}$$

and

$$i = \lfloor x/N_x \rfloor - 1, \tag{2.46}$$

$$j = \lfloor y/N_y \rfloor - 1, \tag{2.47}$$

$$k = \lfloor z/N_z \rfloor - 1, \tag{2.48}$$

$$u = x/N_x - \lfloor x/N_x \rfloor, \tag{2.49}$$

$$v = y/N_y - \lfloor y/N_y \rfloor, \tag{2.50}$$

$$w = z/N_z - \lfloor z/N_z \rfloor. \tag{2.51}$$

One attractive feature of the B-spline model is that it has local support. This means

that changing a control point $d_{i,j,k}$ only affects its local neighborhood, which makes it less computationally expensive even for a large number of control points. The size of the control point grid determines the degrees of freedom of the transformation and thus the difficulty of optimization. In 3D image registration, the B-Spline model can easily have hundreds of parameters. Therefore, choosing suitable optimizers is significant to implement deformable registration with a high dimensional parameter space.

2.5 Shape Representation Using the Poisson Equation

Silhouette contours contains detailed information about the shape of objects and are usually available as segmentation results in computer vision systems (see examples in Fig. 2.4). Much research has been done on using properties extracted from silhouettes. A practical approach to extract properties from silhouettes is to assign to every internal point a value that depends on the relative position of points within the silhouette. For example, the distance transform assigns to every point within the silhouette its minimal distance to the boundary contour. Gorelick *et al.*proposed an approach based on the idea of random walks [13]. If we place a set of particles at the point and let them move randomly until they hit the boundary contour, the expected time for particles hitting the boundaries measures the value of the point.

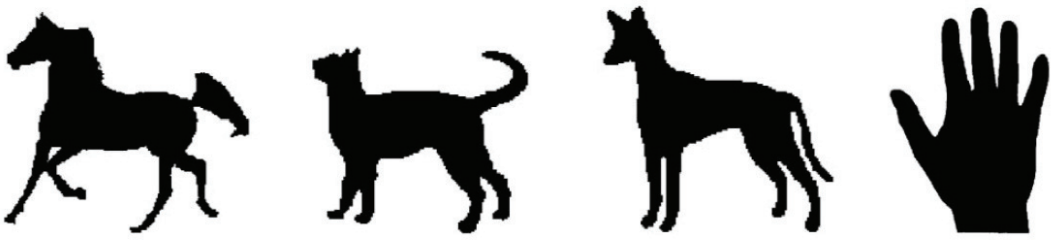


Figure 2.4: A collection of silhouettes [13]

Consider a silhouette S embedded in a grid with mesh size h with a simple closed

contour ∂S. The value of a point is donated as $U(x, y)$. For every point (x, y) within S, $U(x, y)$ equal to the average value of its four nearest neighbors plus a constant representing the time required to get to a neighbor. When we set the constant to one time unit,

$$U(x, y) = 1 + \frac{(U(x + h, y) + U(x - h, y) + U(x, y + h) + U(x, y - h))}{4}. \tag{2.52}$$

Note that (2.52) is a discrete approximation of the Poisson equation

$$\Delta U(x, y) = -\frac{4}{h^2}, \tag{2.53}$$

where $\Delta U(x, y)$ is the Laplacian of U and h is the grid spacing. If we set $\frac{4}{h^2} = 1$, which meaning one spatial unit per time unit, then

$$\Delta U(x, y) = -1, \tag{2.54}$$

where $(x, y) \in S$. Note that on the contour boundary ∂S, $U(x, y) = 0$. Therefore, we are solving the nonhomogeneous Poisson equation (2.54) with Dirichlet boundary condition $U(x, y) = 0$ at contours ∂S.

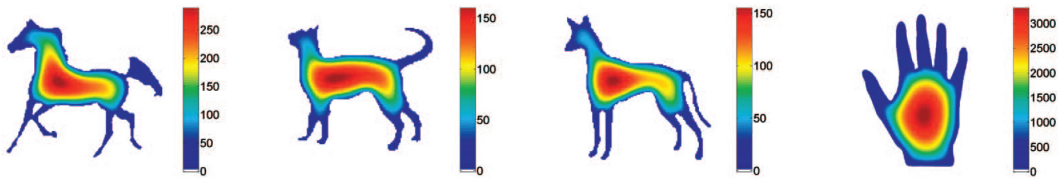


Figure 2.5: Solutions to the Poisson equation for the silhouettes in Fig. 2.4 [13]

Fig. 2.5 shows the solution to the Poisson equation for the silhouettes in Fig. 2.4. We can see that the central part of the shape attains high values of U while the external protrusions has relatively low values of U. A extension of this approach [33] modifies the random walk by adding a an exponential lifetime for the particle and then the expected hitting time satisfies a screened Poisson equation. The modified approach can represent

points both interior and exterior to the shape.

In our method, a 3D version of this shape representation approach will be applied to the cochlea structure to determine the spatial weights for concentration averaging, which will be shown in Section 3.4.

Chapter 3

Methodology

In this chapter, our systematic approach for analyzing μCT images of the mouse cochlea is introduced in detail. An overview is given in section 3.1. The μCT imaging experiments and calibration process are presented in section 3.2. Section 3.3 introduces the 3D subject-atlas image registration technique. The 3D spatial weighting and volume discretization method for extracting and representing concentrations is described in Section 3.4.

3.1 Overview of Our Approach

The diagram of our overall system is shown in Fig. 3.1. The mouse cochlea is imaged by μCT at several time points while ioversol, a contrast agent that simulates a drug, is delivered into the inner ear by a micro-pump. We then use the 3D registration of the μCT image to a labeled atlas in order to segment and identify the intra-cochlea structures in the μCT image. The registration is only performed at the first time point, before the ioversol has reached the inner ear. In a successful registration, the transformation between atlas image and subject μCT image is determined and then applied to label maps of atlas. The transformed label maps of cochlear structures are used to segment and identify each structure in the μCT images. To characterize the diffusion of ioversol in the cochlea, the fluid-filled scalae of the inner ear (scalae tympani (ST), scala vestibuli (SV) and scala

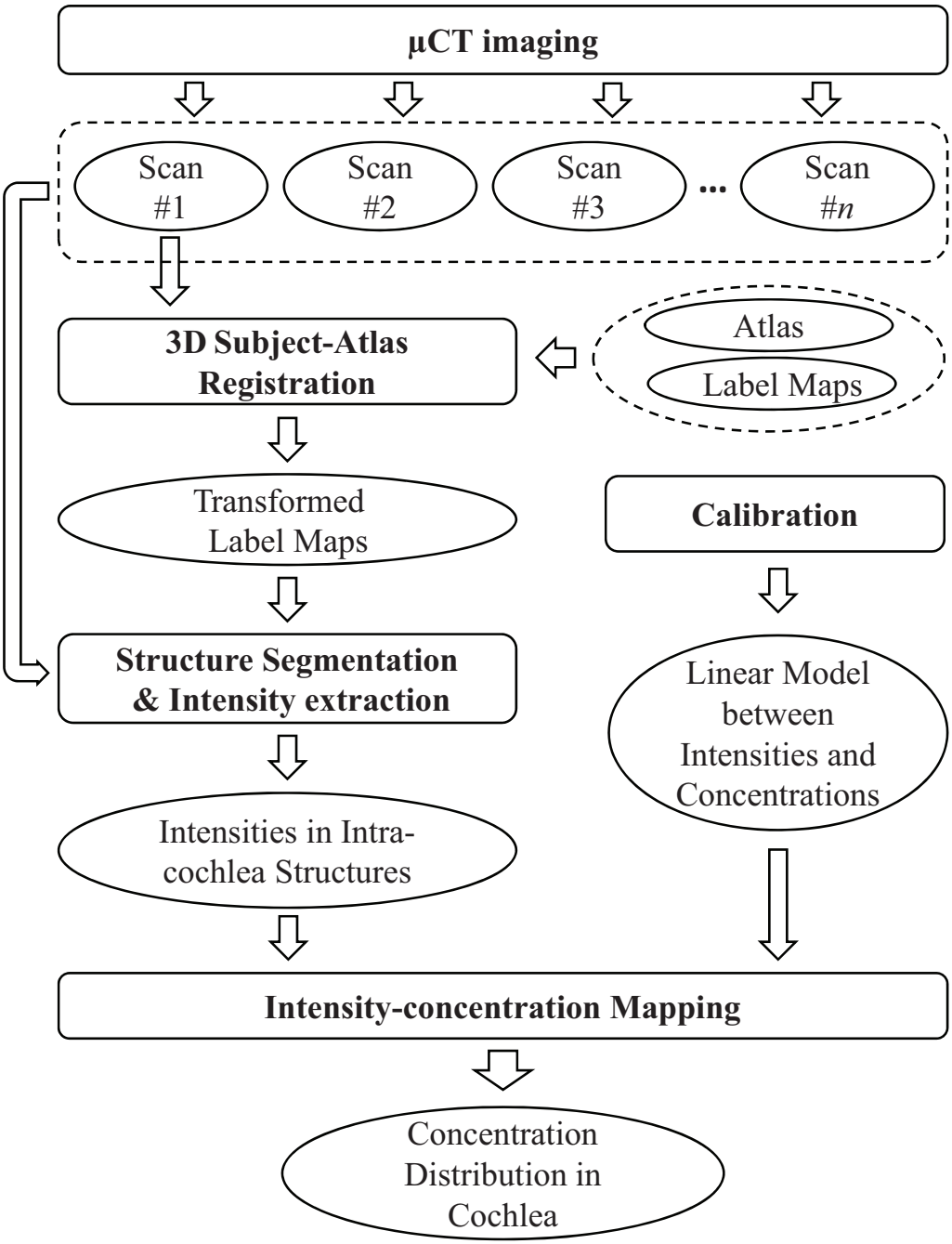


Figure 3.1: Diagram of the method for non-invasive concentration measurement in the mouse cochlea.

media (SM)) and spiral ligament (SL) are segmented in our experiments.

The segmentation results from the first time point are propagated to the subsequent time points since the mouse is fixed in position using a folder. By simple `AND` operations between segmentation results and μCT images, the intensities of each structure are extracted. A prior calibration step enables us to establish a linear model between image intensities in HU and ioversol concentrations. Finally, the extracted intensities are mapped to drug concentration values with the linear model determined by the calibration.

To establish drug pharmacokinentics in cochlea, we further partition each compartment (ST, SV, SM) into several components and perform weighted averaging within each of the partitioned sub-volumes. The weight of a pixel depends on its distance to the boundary of the structure and is computed by solving the Poisson equation introduced in Section 2.5. The concentrations extracted from different series of scans and different components can be used to estimate the spatiotemporal distribution of ioversol in cochlea structures.

We carry out a series of experiments to demonstrate that our technique achieves more accurate registration results compared with Haghpanahi's 2D method, and our approach can measure the drug concentration at any position in the cochlea once it is labeled in the atlas.

3.2 Micro-CT Imaging and Calibration

The animal experiment setting is quite similar to the procedure described in [5]. A customized mouse holder is used for fixing the mouse. This holder minimizes the movement of mouse during the imaging process, which allow us to do subject-atlas registration only for the first scan and apply the segmentation results to the following scans.

We used the same drug delivery system described in [14]. A baseline scan was performed before the pump started to work. Then several series of scans were performed when the contrast agent was delivering into inner ear by a micropump.

Since full-head images are obtained in our experiment, it is necessary to localize the cochlea region from the full head 3D μCT image prior to image registration, as shown in Figure 3.2. We performed this manually; however, it could be done automatically, for

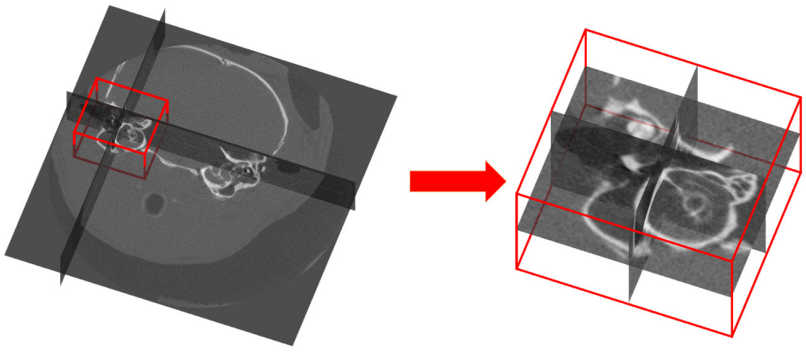
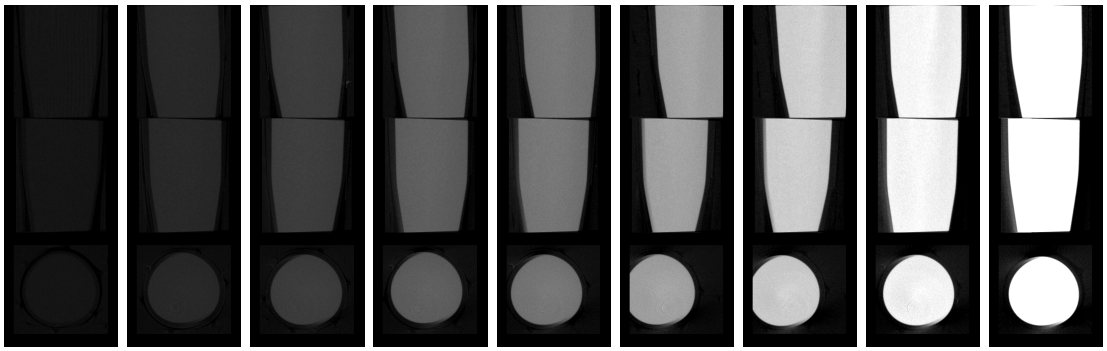


Figure 3.2: Extract the region of cochlea (red box) from full head micro-CT scans

example, by searching for geometric features matching those of the atlas that are preserved when reflected through the sagittal plane.



(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1 g/ml saline	18.98 mg/ml	40.68 mg/ml	75.95 mg/ml	100.34 mg/ml	149.16 mg/ml	200.69 mg/ml	249.51 mg/ml	320 mg/ml

Figure 3.3: Calibration images. Three orthogonal 2D views (in rows) of nine calibration 3D image volumes (in columns). The first sample is 1g/ml saline; the 2nd to 9th samples are ioversol resolution of 18.98mg/ml, 40.68mg/ml, 75.95mg/ml, 100.34mg/ml, 149.16mg/ml, 200.69mg/ml, 249.51mg/ml and 320mg/ml equivalent iodine concentration.

The pixel intensities of μCT images are quantified in *Hounsfield Units*(HU), which is a quantitative scale of radio-density. HU values are calculated from the original linear attenuation coefficient by a transformation so that the radio-density of distilled water at standard pressure and temperature (STP) is defined as zero HU, while the radio-density

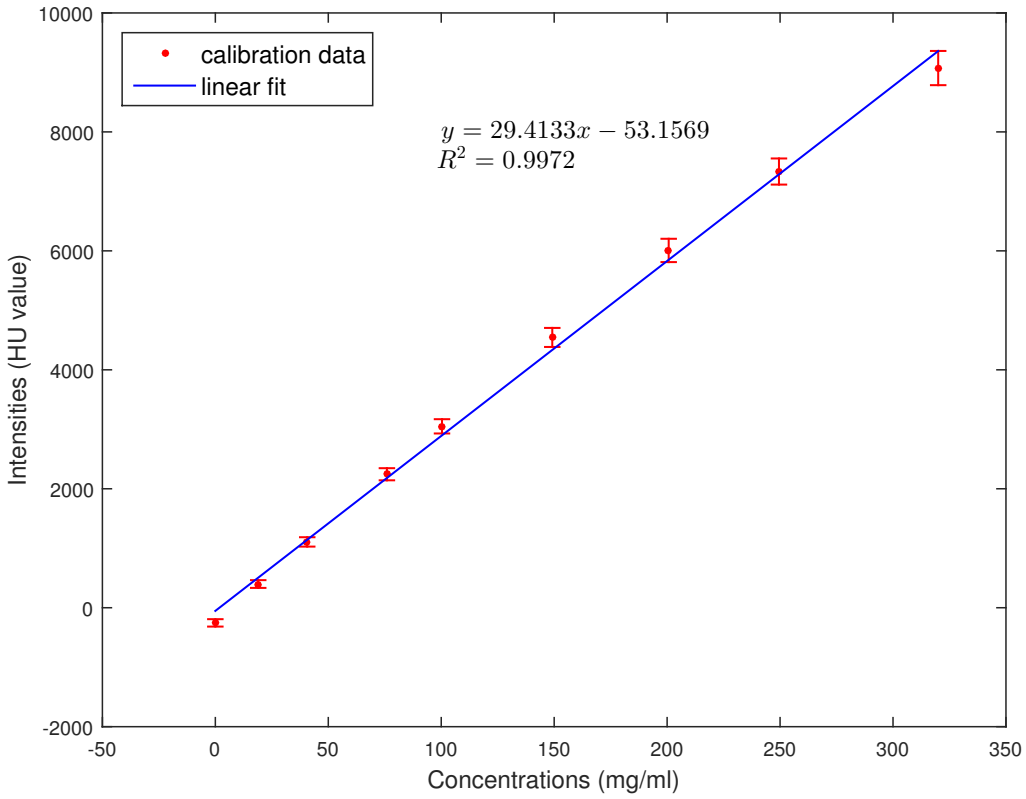


Figure 3.4: The calibration curve of ioversol concentration and image intensities. Red dots are real data points and the blue line is the lowest least-squares fit of the data to a line.

of air at STP is defined as -1000 HU. The transformation is formulated as:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}, \tag{3.1}$$

where μ_{water} and μ_{air} are the linear attenuation coefficients of water and air respectively.

To build the correlation between the image voxel intensity and drug concentration, ioversol solutions of multiple concentrations are imaged in the same micro-CT system. The scans are shown in Fig. 3.3. A 1g/ml saline solution and ioversal solutions of 18.98mg/ml, 40.68mg/ml, 75.95mg/ml, 100.34mg/ml, 149.16mg/ml, 200.69mg/ml, 249.51mg/ml and 320mg/ml equivalent iodine concentration are imaged. Segmentation is done in `ITKSNAP` for the 3D images and the average intensity of each solution is extracted. Using linear regression, the mapping between image intensities and ioversol concentrations is established, as shown in Fig. 3.4. The red points are the calibration data points and the error bar is the estimated standard deviation of sample HU values in calibration images for each concentration. The blue line is the linear fitting result.

The best-fit linear model is

$$y = 29.4133x - 53.1569, \tag{3.2}$$

where y represents the HU value and x represents the equivalent iodine concentration. The inverse function of (3.2) will be used to transform the image intensities to ioversol concentrations.

We can observe that the estimated standard deviation increases as the concentration/HU values increase. This phenomenon is caused by the shot noise in x-ray tomography. Shot noise follows Poisson distribution and its standard deviation is linearly correlated with the square root of the number of photons/electrons, meaning that the higher image intensities yield higher δ.

To address the accuracy of our calibration results, the confidence intervals of the HU

	Mean	Upper	Lower
1	0.0000	0.0007	-0.0007
2	18.9800	18.9819	18.9781
3	40.6800	40.6823	40.6777
4	75.9530	75.9560	75.9500
5	100.3400	100.3435	100.3365
6	149.1600	149.1645	149.1555
7	200.6900	200.6955	200.6845
8	249.5100	249.5165	249.5035
9	320.0000	320.0079	319.9921

Table 3.1: Upper and lower bounding values of 95% confidence interval for calibration concentrations (mg/ml)

sample means for each concentration are calculated according to:

$$[\bar{Y} - z\frac{\delta}{n}, \bar{Y} + z\frac{\delta}{n}], \tag{3.3}$$

where $\bar{Y}$ is the sample mean of HU values and δ is the estimated standard deviation of sample HU values, shown as red points with error bar in Fig. 3.4. z is the upper (1-C)/2 critical value for the standard normal distribution where C is the confidence level. We have confidence C that the real mean of HU distribution is within the interval. For each calibration data point, there are upper and lower bounds for the confidence interval. Two bounding lines can be estimated from the upper and lower bound points, respectively, that represent the accuracy range of the HU-concentration transformation. At a confidence level of 95% where $z = 1.96$, the estimated bounding lines are:

$$\text{Upper bounding line:} \qquad y^U = 29.4138x - 53.1131 \tag{3.4}$$

$$\text{Lower bounding line:} \qquad y^L = 29.4127x - 53.2007. \tag{3.5}$$

According to the bounding lines, the upper and lower bounds for concentration values in the calibration data are given in Table 3.1.

3.3 3D Subject-atlas Image Registration

To segment and classify cochlea from μCT images, the atlas based segmentation method is used. The process is described in Fig. 3.5. Firstly, the 3D subject-atlas registration searches for the optimal transformation T^* between the atlas (top left) and μCT images (top right). Then T^* is applied to the label map of the atlas (bottom left), and the transformed label will automatically segment the μCT (bottom right). To measure concentrations of ioversol in cochlear scalae, SV (blue), SM (yellow), ST (red) are segmented. SL (violet) also allows intra-scalae drug transport and thus it is segmented as well.

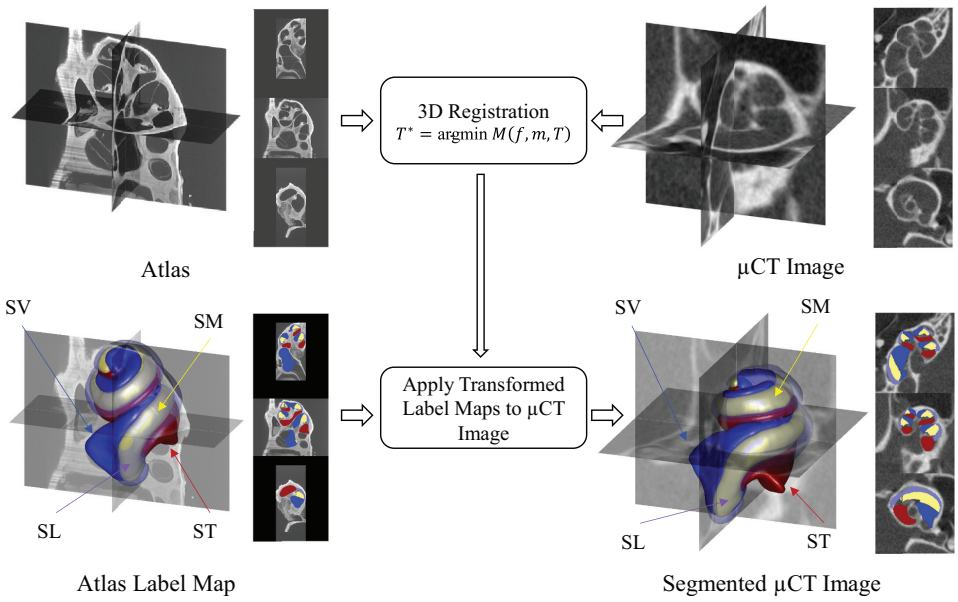


Figure 3.5: Registration between 3D labeled atlas and μCT image of mice cochlea. Top left: Atlas of mouse cochlea; Top right: μCT image of cochlea to be segmented. Bottom left: Atlas and its label maps of SV (blue), SM (yellow), ST (red), and SL (violet); Bottom right: Segmented μCT image of cochlea by transformed label maps.

In our 3D subject-atlas registration method, the mouse cochlea database (MCD) of Santi et al. [37] is used as the atlas. The MCD includes a stack of 2D sectional images of a mouse inner ear that has been mannually segmented and labeled. Among the structures

labeled in the MCD, only scala tympani (ST), scala vestibuli (SV), scala media (SM) and spiral ligament (SL) are used.

In the registration framework given in Fig. 2.1, the atlas image of higher resolution works as the moving image and the μCT image of lower resolution works as the fixed image. One reason is that the moving image is interpolated into the fixed image space and interpolating a lower resolution image to a higher resolution does not actually produce additional information while it increases the computational cost of registration. Additionally, it is also convenient to keep the μCT images unchanged to extract raw pixel intensities from them.

We implement the registration algorithm in Insight Segmentation and Registration Toolkit (ITK). Multi-stage and multi-resolution strategies are used to increase the speed and robustness of our registration algorithm. These are introduced in the next section.

3.3.1 Advanced Strategies of Image Registration

To achieve higher registration speed, accuracy and robustness, multi-resolution and multi-stage strategies are used in our registration algorithm. The basic idea of multi-resolution registration is that the registration is done starting from a coarse scale where the images have lower resolution. The optimal transformation determined by the coarse level registration is used to initialize the registration at the next finer scale. The process is repeated until the finest resolution is reached. Fig. 3.6 presents an example of three-level multi-resolution registration. Note that any number of resolution levels can be used with this strategy. This strategy helps to eliminate the possibility of being trapped in a local minimum and thus increases the registration success rate and robustness. Additional smoothing at coarse levels can help even more. Since fewer optimization iterations are run at the full-resolution level, the multi-resolution strategy can speed up the registration process greatly, especially for a large dataset like the high resolution 3D image in our case.

The other scheme known as multi-stage starts the registration from a simple transform, e.g. a similarity transformation, to a more complicated transformation, e.g., an affine transformation, and finally to a deformable transformation. The output of registration

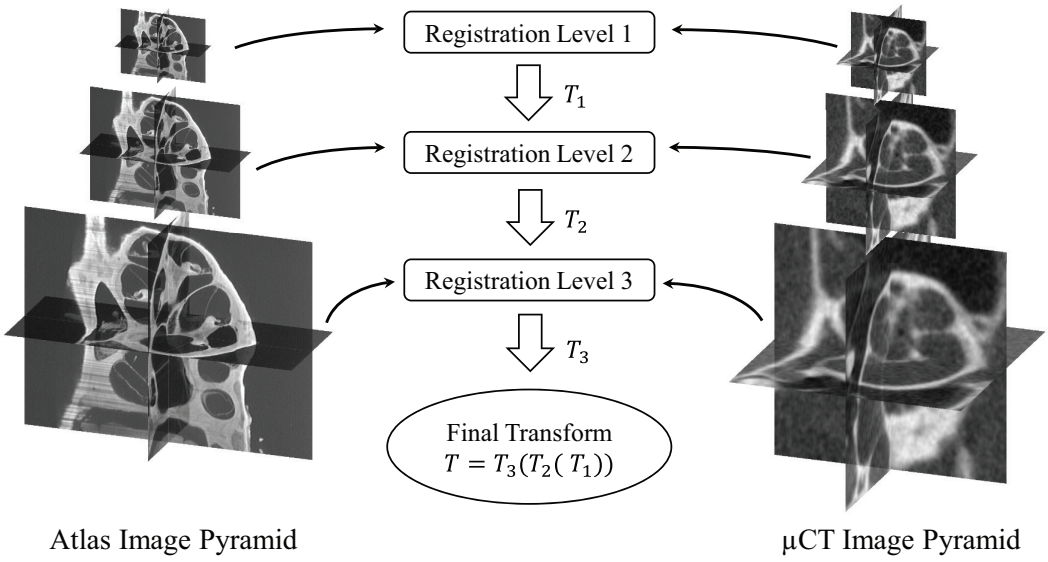


Figure 3.6: Multi-resolution strategy for image registration is to register the moving image (atlas) pyramid to the fixed image (micro-CT scans) pyramid. The output of lower level registration will works as the initial transform of higher level registration. The higher level registration utilizes the higher resolution image in the image pyramid.

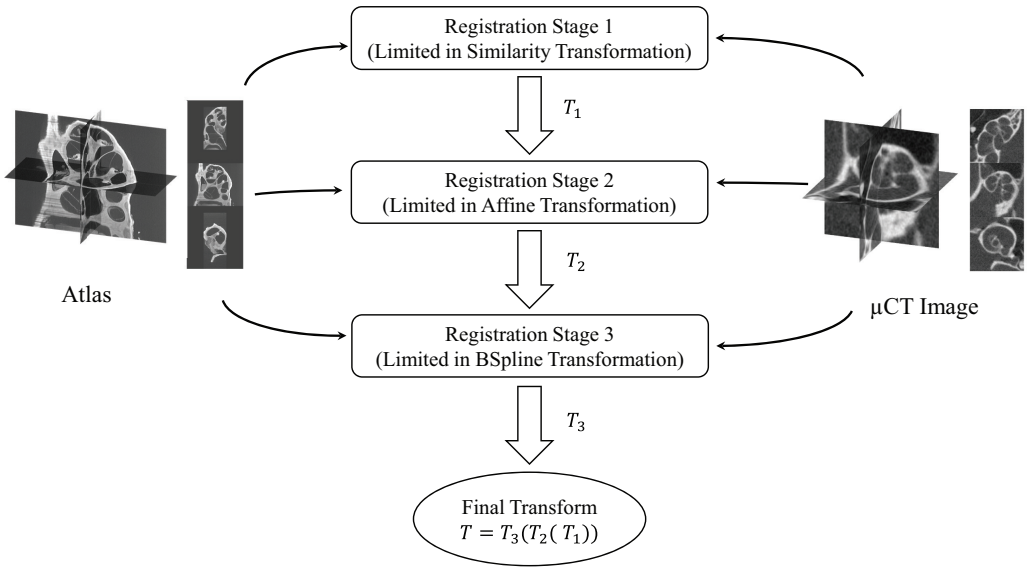


Figure 3.7: Multi-stage strategy for image registration is to register the moving image (atlas) to the fixed image (micro-CT scans) with different transformation limitation. The output of lower level registration will works as the initial transform of higher level registration. The higher level registration uses the transformation with the higher degree of freedom.

with a simple transformation is used as the initial transformation of the next registration stage with the more complicated transformation. Similar to the multi-resolution strategy, the multi-stage strategy could improve the robustness and success rate of registration. It also saves a great amount of computation time because the deformable registration starts when we are already close to the optimal value.

A combination of multi-resolution and multi-stage strategies is recommended. In each stage of registration, we use different resolution level or multiple resolutions. For example, the images that shrinked with the factor 3 and 2 are respectively applied to similarity and affine stages, and then the image with original resolution is used for the deformable registration.

3.3.2 Image Registration in ITK

Our registration algorithm is implemented using the Insight Segmentation and Registration Toolkit (ITK). ITK is an open-source, object-oriented and cross-platform software system for image processing, segmentation, and registration. It is popular in the medical image analysis community.

The various image input/output (IO) types and filters make it convenient to read, write and process medical images of different formats. For example, the μCT scan series of 2D DICOM images can be easily read and writen as a 3D image file in the NIFTI format.

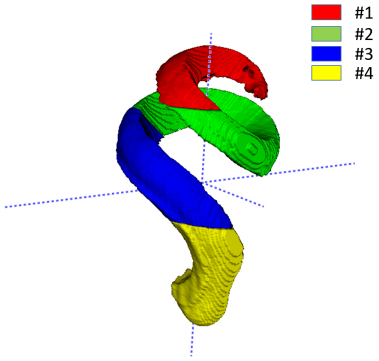
ITK provides an extensible registration framework of pluggable components that can be easily interchanged; e.g., similarity metrics, optimizers, transformation type and interpolation methods. It should be noted that, in ITKv4, the registration is done in physical coordinates and the computation happens on a physical grid, called the virtual image, rather than the fixed image domain [17]. This framework allows us to register two images with different scales, sizes, and resolution easily, and we can also write the output image with any desired size and resolution. However, it requires the correct image origin, pixel size, and image direction for successful registration.

We can also easily handle the multi-resolution and multi-stage registration at the same time. ITK allows us to set a registration filter with different types of transformations and multiple resolution levels by using the defined *ShrinkFactor* and *SmoothingSigma* vectors.

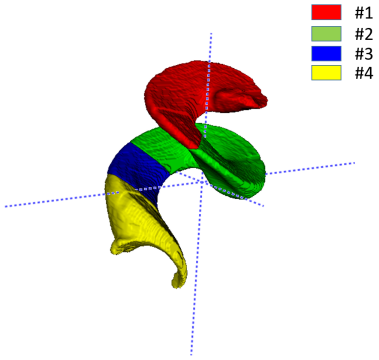
3.4 Spatial Weighting and Volume Discretization

In order to analyze the drug distribution over the cochlea, each cochlear structure is discretized. It means one structure, like SV, is divided into several sub-structures. The ST and SM are divided into four parts and the SV is divided into five parts. The discretization is shown in Fig. 3.8.

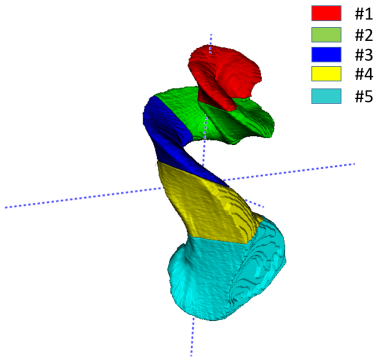
The image intensities, or ioversol concentrations after mapping, will be integrated over each sub-region to compute a weighted average. The weight of each pixel depends on the distance to the boundary of the structure. Pixels that are closer to the boundary



(a) SM substructures

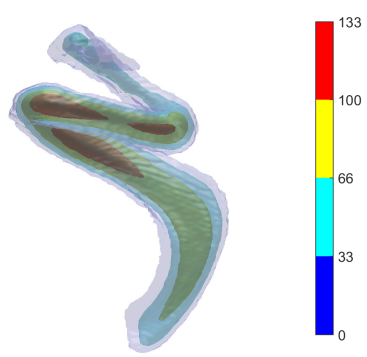


(b) ST substructures

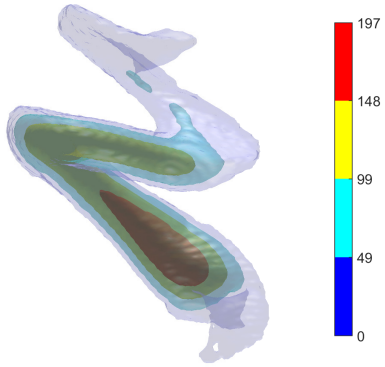


(c) SV substructures

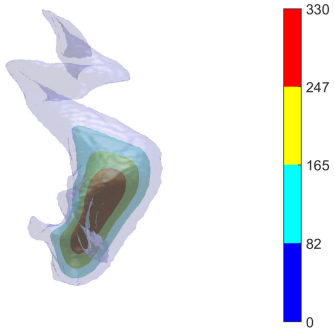
Figure 3.8: Discretization of intra-cochlea structures. For the convenience of concentration quantification, ST, SV, SM are divided into several component, represented by different colors. The component numbers are increasingly ordered from top to bottom.



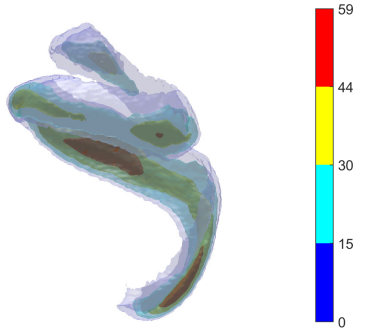
(a) Weight volume of SM



(b) Weight volume of ST



(c) Weight volume of SV



(d) Weight volume of SL

Figure 3.9: Weights of intra-cochlea structures. By solving a non-homogeneous Poisson equation, the weights representing the distance to structure boundary is determined for weighted concentration averaging.

are assigned lower weights. Recalling the shape representation method by solving a Poisson equation in section 2.5, the weights of ST, SV and SM are computed. A multi-grid algorithm [7] implemented in MATLAB [12] is used to solve the Poisson equation. A 3D rendering of the weights is shown in Fig. 3.9. We can see that higher weights are concentrated in the center structures while weights on the boundary are zero.

Using the spatial weights and volume components, several observation points are obtained instead of all pixel positions so that the drug spatial distribution is described and observed succinctly. However, we can still measure the drug concentration at a resolution as high as that of the μCT image.

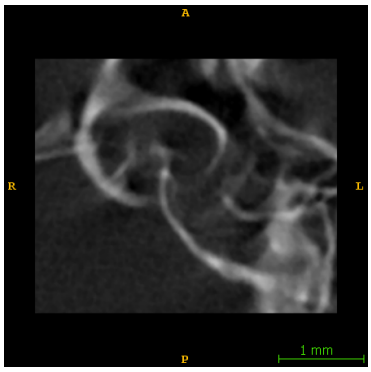
Chapter 4

Experimental Results

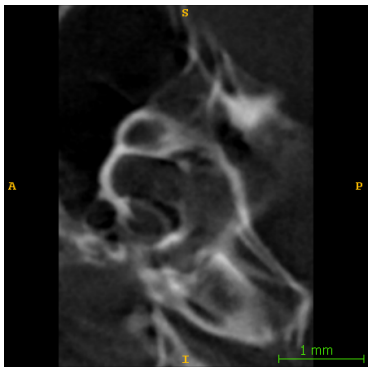
4.1 Data for Experiments

Our approach is applied to a μCT image series of mouse cochlea to illustrate its effectiveness. The scan #0 was obtained before the micro-pump started to deliver the contrast agent solution into the mouse ear. Additional four scans are performed at 30 min, 1 hour, 1 hour and 50 minutes, and 2 hours and 40 minutes after the first scan. The three orthogonal 2D views of each scan are shown in Fig. 4.1. Note that they are manually extracted from the full-head scans. The size of the cochlear image volume from each scan is $217 \times 183 \times 260$ and the voxel spacing is $0.01638\ mm \times 0.01638\ mm \times 0.01638\ mm$ in the x, y, and z directions respectively.

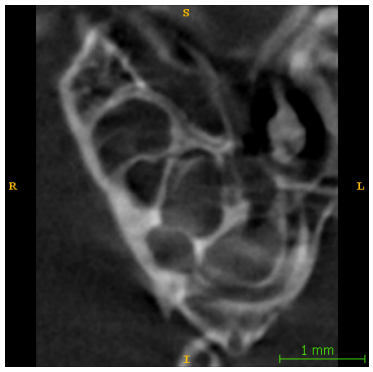
The mouse cochlea atlas introduced in Section 3.3 is a stack of 69 2D images of size 632×480, spacing $0.005435\ mm \times 0.005435\ mm$ and slice thickness $0.02mm$. The label of intra-cochlea structures is originally represented as contours labeled with different colors on the atlas images. We extract the contours and fill them into label maps for the convenience of segmentation operations. The atlas and label maps of ST, SV, SM and ST are shown in the bottom left of Fig. 3.5.



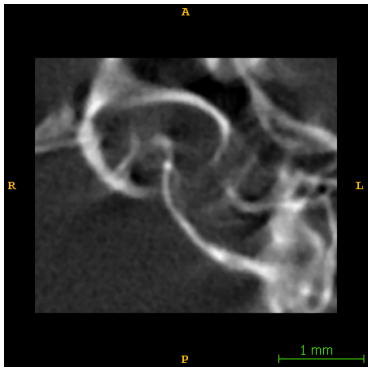
(a) Axial view of scan #0



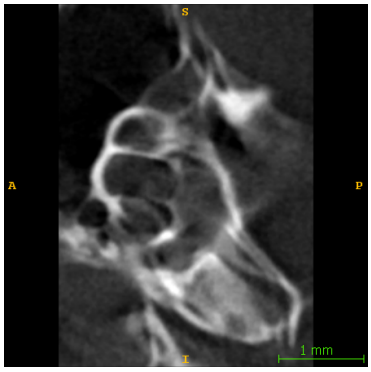
(b) Sagittal view of scan #0



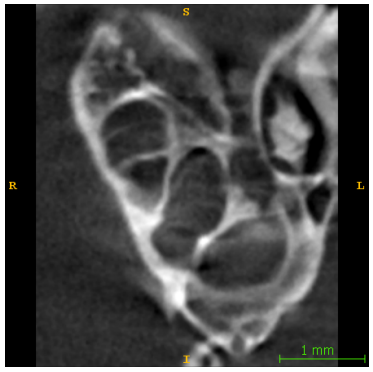
(c) Coronal view of scan #0



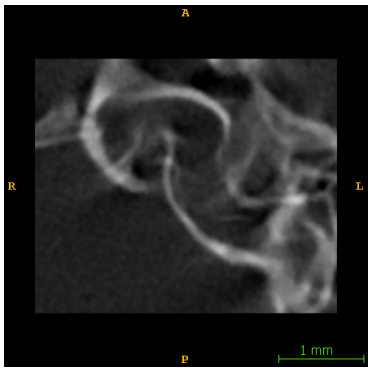
(d) Axial view of scan #1



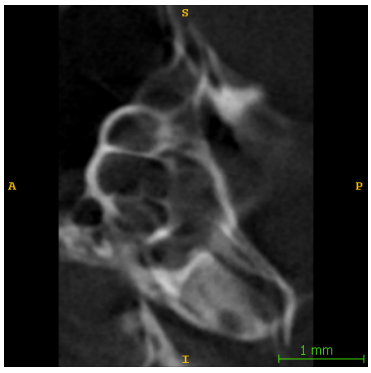
(e) Sagittal view of scan #1



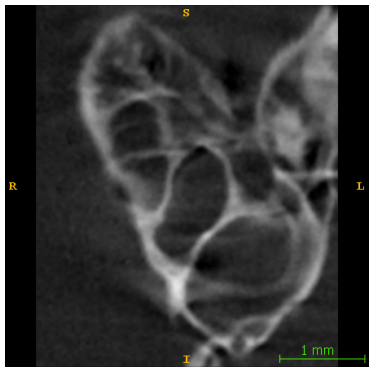
(f) Coronal view of scan #1



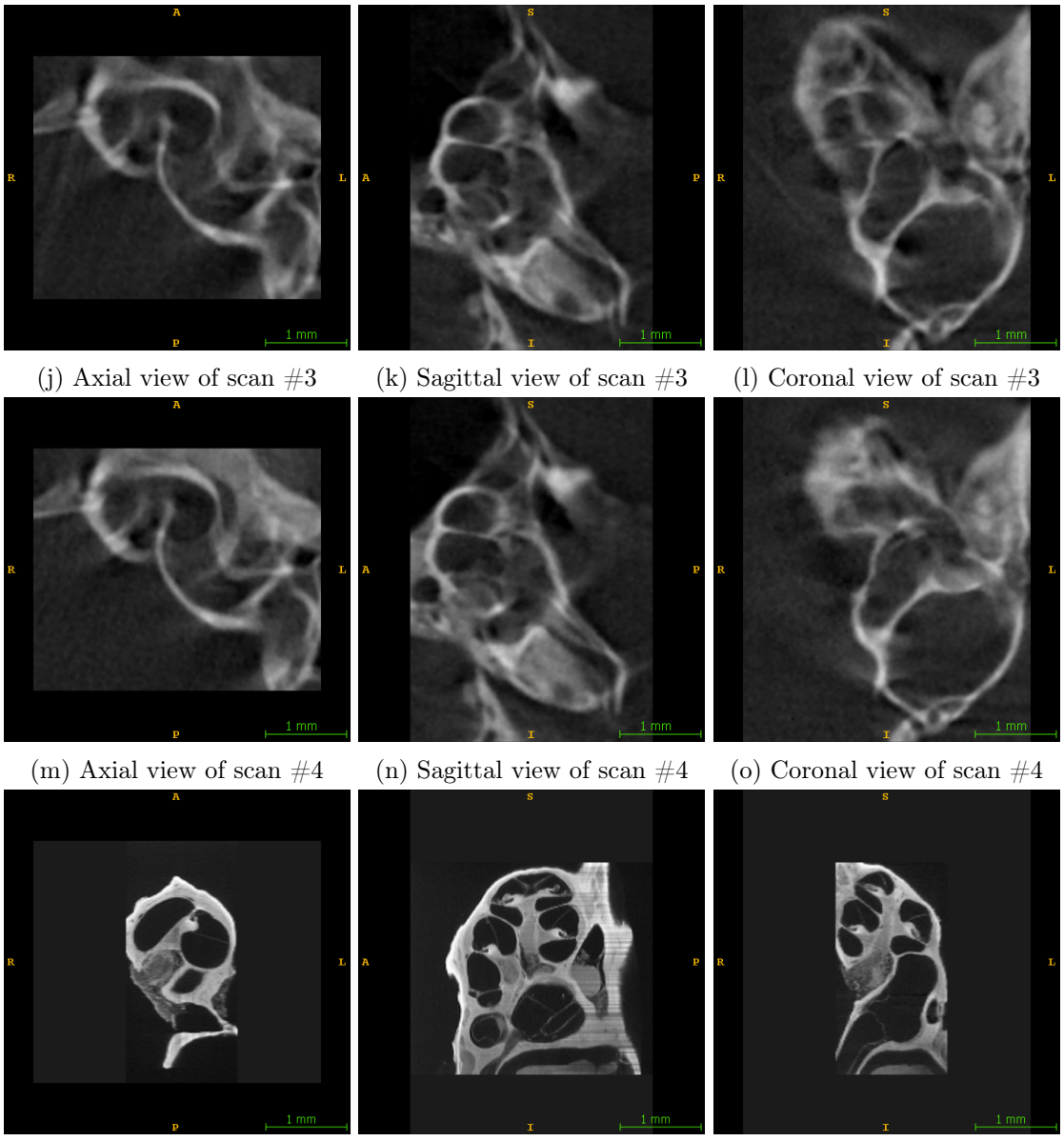
(g) Axial view of scan #2



(h) Sagittal view of scan #2



(i) Coronal view of scan #2



(j) Axial view of scan #3 (k) Sagittal view of scan #3 (l) Coronal view of scan #3

(m) Axial view of scan #4 (n) Sagittal view of scan #4 (o) Coronal view of scan #4

(p) Axial view of atlas (q) Sagittal view of atlas (r) Coronal view of atlas

Figure 4.1: Three orthogonal 2D views of μCT scans of mouse cochlea during ioversol delivery and atlas. Different scans are represented in different rows and atlas is in the last row. The left column are axial views, the center column are sagittal views and the right column are coronal views.

4.2 Experiments: Registration

The registration framework we proposed is flexible and can be easily customized for different datasets and requirements. In our case, the composite of three registration stages is applied with various resolution levels, shown in Table 4.1. Our registration algorithm includes three stages that are respectively limited to similarity, affine and B-Spline transformation. The similarity stage uses two resolution levels and the other two stages use only one. All stages use the*Regular Step Gradient Descent Optimizer* in ITK which is a variant of gradient descent that reduces the step length by a relaxation factor once the direction of the derivative changes. Normalized Correlation Coefficient works as the similarity metric for all stages.

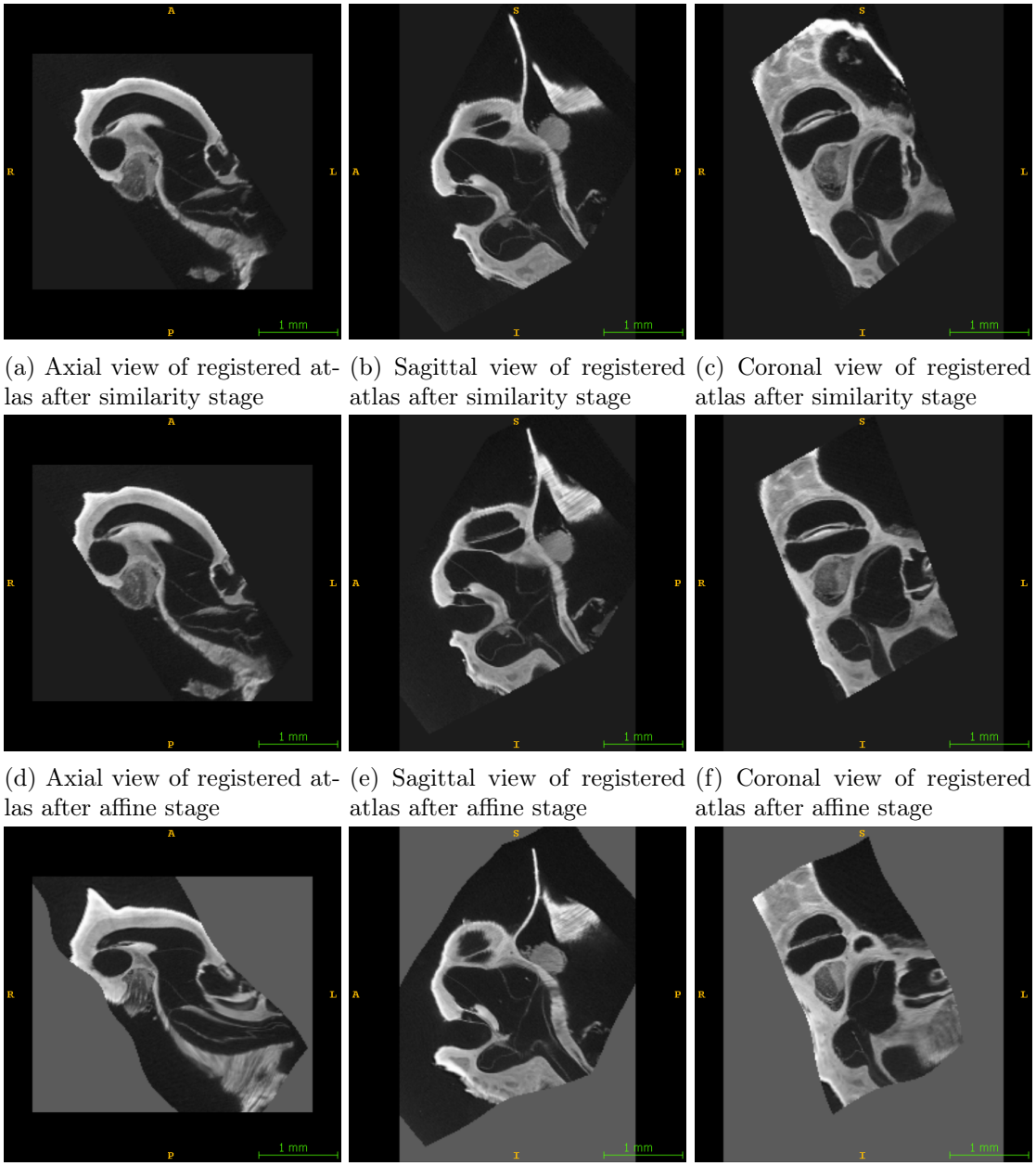
The optimization parameter for registration need to be tuned individually for each registration case. The working parameters are mostly found by trial-and-error, but the analysis for each case is also important. Some guidelines can be found in [17].

Stage#	**1**	**2**	**3**
Transformation	Similarity	Affine	Deformable (B-Spline) Grid Node # in one dimension = 10
Resolution Level (Shrink Factor)	3 2	1	2
Optimizer	Regular Step Gradient Descent Optimizer		
Similarity Metric	Normalized Correlation Coefficient		

Table 4.1: Setting for registration.

As we mentioned in section 3.3.1, the output transformations of earlier registration stages work as initial transformations for following stages. Before the first similarity stage, the moments of gray level values of both images are calculated and the center of mass of the moving image is used as the rotation center. The translation between the two centers of mass is used to initialize the first registration stage.

The registration results of each stage are shown in Fig. 4.2, and their overlapping with the μCT image is shown in Fig. 4.3. These figures show that the similarity and affine transformations roughly align the atlas with the μCT image by rotation, translation and



(a) Axial view of registered atlas after similarity stage

(b) Sagittal view of registered atlas after similarity stage

(c) Coronal view of registered atlas after similarity stage

(d) Axial view of registered atlas after affine stage

(e) Sagittal view of registered atlas after affine stage

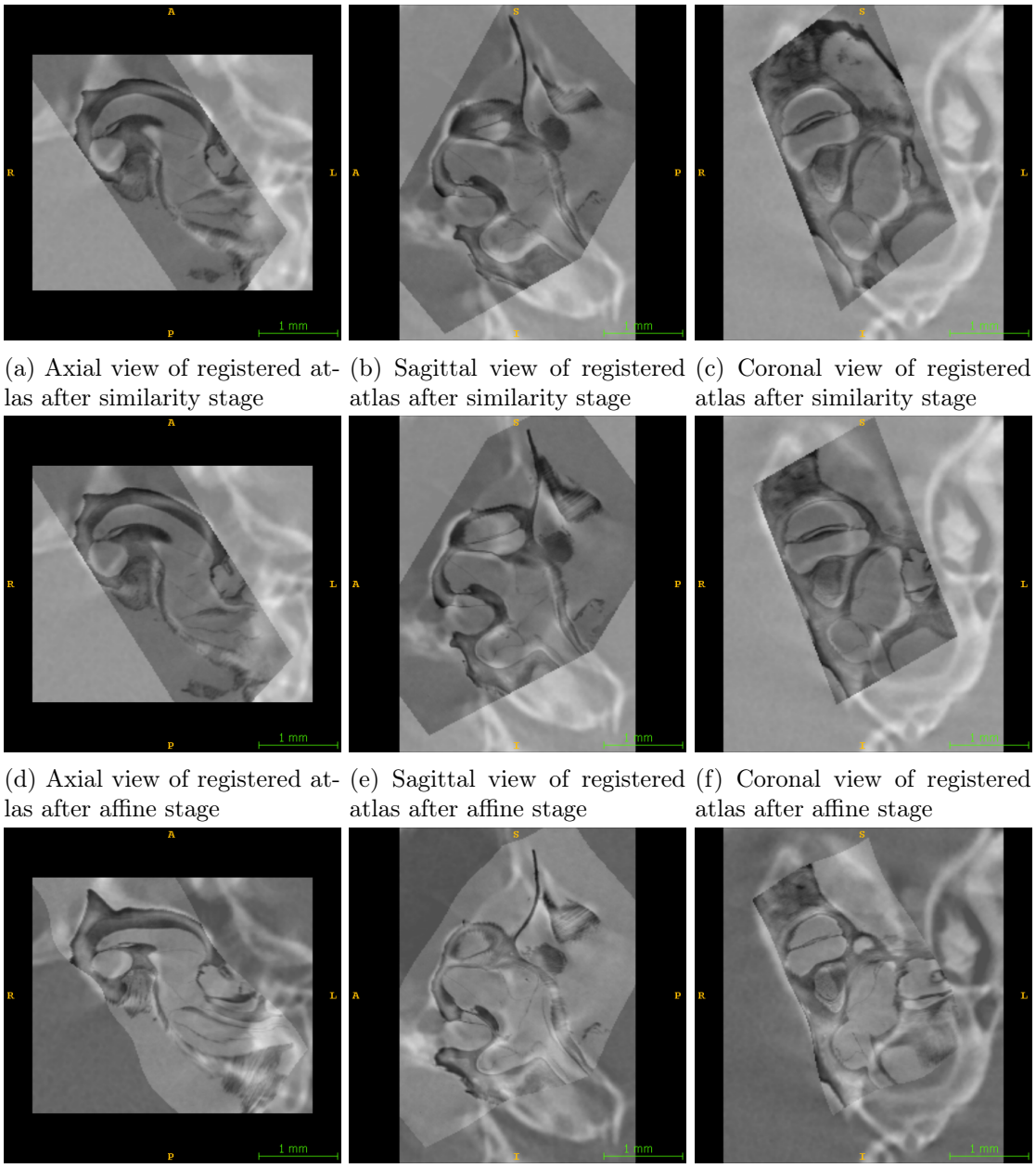
(f) Coronal view of registered atlas after affine stage

(g) Axial view of registered atlas after B-Spline stage

(h) Sagittal view of registered atlas after B-Spline stage

(i) Coronal view of registered atlas after B-Spline stage

Figure 4.2: Three orthogonal 2D views of registered atlas image after each stage. Different stages are represented in different rows. The left column are axial views, the center column are sagittal views and the right column are coronal views.



(a) Axial view of registered atlas after similarity stage

(b) Sagittal view of registered atlas after similarity stage

(c) Coronal view of registered atlas after similarity stage

(d) Axial view of registered atlas after affine stage

(e) Sagittal view of registered atlas after affine stage

(f) Coronal view of registered atlas after affine stage

(g) Axial view of registered atlas after B-Spline stage

(h) Sagittal view of registered atlas after B-Spline stage

(i) Coronal view of registered atlas after B-Spline stage

Figure 4.3: Three orthogonal 2D views of the overlapping between registered atlas image of each stage and the μCT images of scan #1. Different stages are represented in different rows. The left column are axial views, the center column are sagittal views and the right column are coronal views.

rescaling, and the deformable registration transforms the atlas locally to better match the μCT image. The value of the similarity metric at the end of each registration stage is shown in Table 4.2. Since the NCC value is closer to 1 than 0 when the two images are better aligned, we can see each stage helps to improve the similarity between the transformed atlas and the μCT image. Note that in ITK, the value of NCC is negative from 0 to -1 so that the optimizer can *minimize* it.

Registration Stage	Similarity Stage	Affine Stage	Deformable Stage
NCC Value	0.243	0.295	0.467

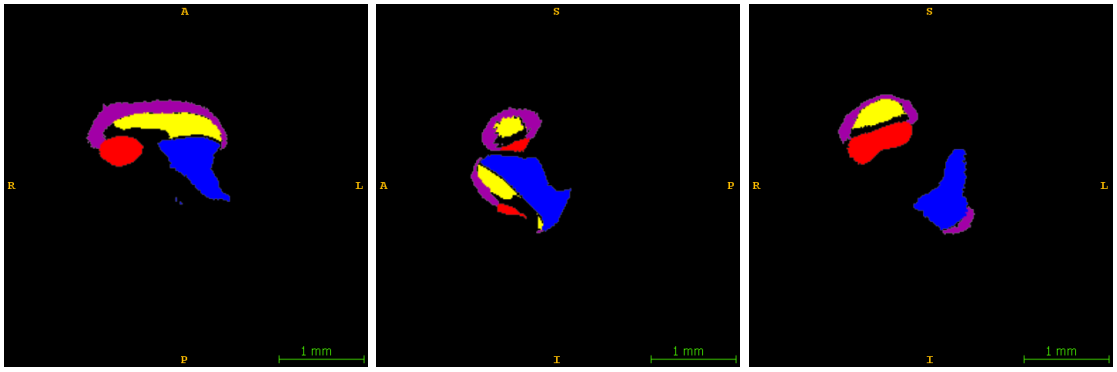
Table 4.2: Value of similarity metric at the end of each registration stage.

Recall that we assume that the mouse head is fixed perfectly and there are no movements during imaging. So only the μCT image of the first scan is registered with the atlas, and its segmentation result is propagated to the next scans. However, in the experiments, we still find that there are on the order of tens of pixels of translation between images of two adjacent scans. This can be solved by adding a easy registration step to find the translations between the first scan and the following scans and composite them with the transformation found from the registration between the atlas and the first scan. This can be formulated as:

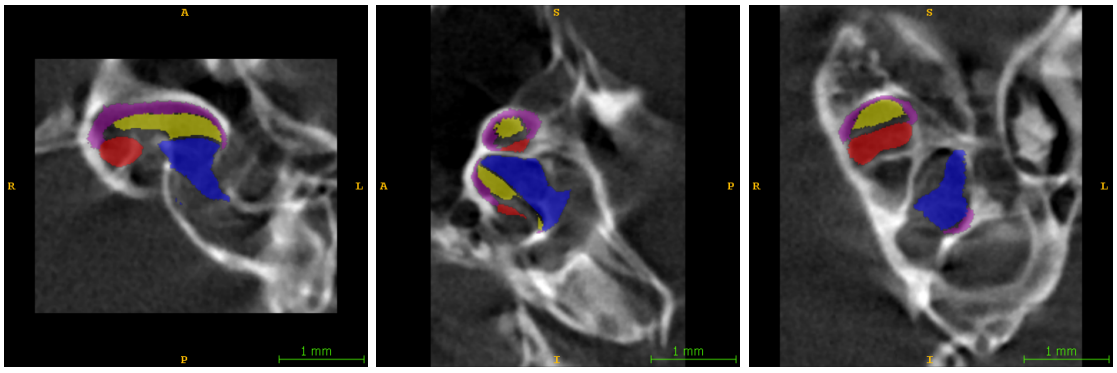
$$T_i^A = T_i^1(T_1^A), \tag{4.1}$$

where T_i^A is the transformation from the atlas to the μCT image of ith scan, and T_i^1 is the transformation from the μCT image of the ith scan to that of the first scan. In this way, the translations between scans are compensated and each scan has a unique segmentation result.

Applying the transformation determined by the three registration stages to label maps enables the segmentation to be carried out automatically. We show the transformed label maps and segmentation results for scan #1 in Fig. 4.4. It can be seen that an accurate segmentation of the intra-cochlea structures have been achieved.



(a) Axial view of transformed label map (b) Sagittal view of transformed label map (c) Coronal view of transformed label map



(d) Axial view of labeled μCT image of scan #1 (e) Sagittal view of labeled μCT image of scan #1 (f) Coronal view of labeled μCT image of scan #1

Figure 4.4: Three orthogonal 2D views of the transformed label map using the transformation output of B-Spline registration stage (the first row) and the labeled μCT images of scan #1 (the second row). The left column are axial views, the center column are sagittal views and the right column are coronal views.

4.3 Experiments: Concentration Extraction and Weighted Averaging

With the segmentation result (transformed label maps), the intensities of each intra-cochlea structure can be easily extracted by an AND operation between label maps and μCT images. With the linear model between intensities and concentrations shown in Fig. 3.4, the concentrations are extracted from the μCT images. The 3D concentration distributions in each cochlear structure in scan #1 are shown as Fig. 4.5.

In Fig. 4.5, we can see that the concentration variation is high and the elevated concentrations around volume boundaries are inaccurately measured. The imaging noise accounts for the negative concentrations which can be reduced by smoothing. The reason of elevated concentrations near the boundary is that because the scalae are tightly surrounded by bone structures, any tiny misplacement of label maps will lead to the wrong intensity extraction from bone structures which have high intensities.

To reduce the effect of incorrectly elevated concentrations on volume boundaries, the weights that depend on the distance to boundary (the closer the distance is, the lower the weights are) are used, which means that the data near the centers of the region get more attention. The weighted concentration distribution is computed by:

$$C_w(x,y,z) = \frac{w(x,y,z)}{\max(w(x,y,z) \in \mathcal{L}_i)} \cdot \mathcal{F}\{I(x,y,z)\}, \tag{4.2}$$

where w is the weight calculated by solving the Poisson equation (2.54) shown in Fig. 3.9, $\mathcal{L}_i$ is the ith label map, and $\mathcal{F}$ is the inverse function of 3.2 that maps the intensities $I(x,y,z)$ into concentrations.

The 3D distribution of weighted concentrations in cochlear structures are shown in Fig. 4.6. We can see that after the weights are applied to the concentrations, those wrongly extracted concentrations from bone structures are mostly compensated.

To quantify the spatial and temporal concentration distribution, the volume discretization is performed as described in Fig. 3.8. The ST, SV and SM compartments are divided into four (SM, ST) or five (SV) components, and the weighted average of concentrations

is computed for each component by:

$$\bar{C}_w = \frac{\sum_\Omega w \cdot C}{\sum_\Omega w}, \tag{4.3}$$

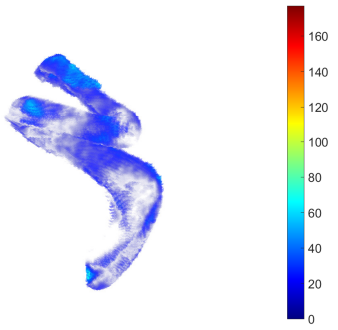
where Ω is the volume of a component, w are weights and $C = \mathcal{F}\{I(x, y, z)\}$ are concentrations in Ω.

The weighted concentration of each component from the baseline scan (#0) is computed and shown in Table 4.3. The 95% confidence interval is also given for each concentration value according to the confidence interval fitted line 3.4 and 3.5 of calibration data. The confidence interval represents the accuracy of our HU value to concentration transformation. The confidence interval is narrow enough to give us accurate estimation and distinction between components.

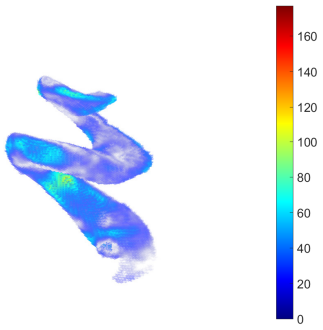
The weighted average concentrations are plotted as spatial and temporal distributions in Fig. 4.7. The concentrations of each components from the baseline scan (#0) have been subtracted from following scans for the convenience of observing the changes of concentrations. Note that the component #1 is the one that is farthest to the round window where the ioversol solution is delivered into cochlea, and the last component is the one that is closest. For the temporal distribution of each structure, no obvious changes of concentrations happen; for the spatial distribution, there is a higher concentration difference in different components of each structure but still no entries of ioversol solution.

Table 4.3: The mean concentrations (mg/ml) with 95% confidence interval ofbaseline scan (#0) in structure SM, ST and SV

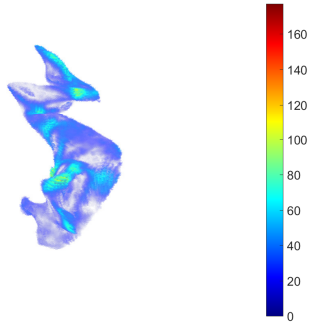
Structures / Component #	SM	ST	SV
1	28.6348 ± 0.0023	20.8543 ± 0.0019	42.6141 ± 0.0023
2	8.2254 ± 0.0019	13.7510 ± 0.0018	20.8045 ± 0.0019
3	14.3630 ± 0.0020	26.4995 ± 0.0020	25.2302 ± 0.0020
4	4.3067 ± 0.0017	23.1823 ± 0.0020	9.6413 ± 0.0017
5	–	–	9.6413 ± 0.0017



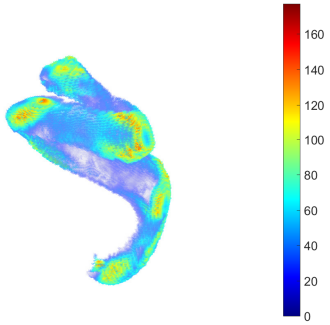
(a) Concentration in SM from scan# 1



(b) Concentration in ST from scan# 1

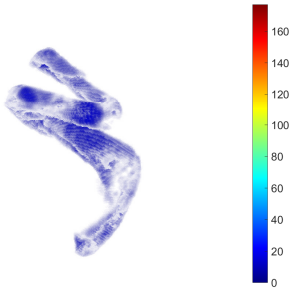


(c) Concentration in SV from scan# 1

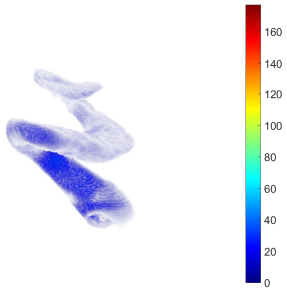


(d) Concentrations in SL from scan# 1

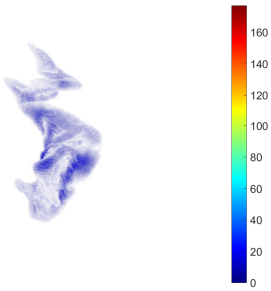
Figure 4.5: 3D rendering of unweighted concentration (mg/ml) from scan #1



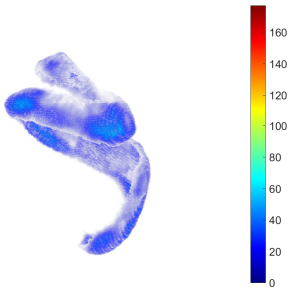
(a) Weighted concentration in SM from scan# 1



(b) Weighted concentration in ST from scan# 1

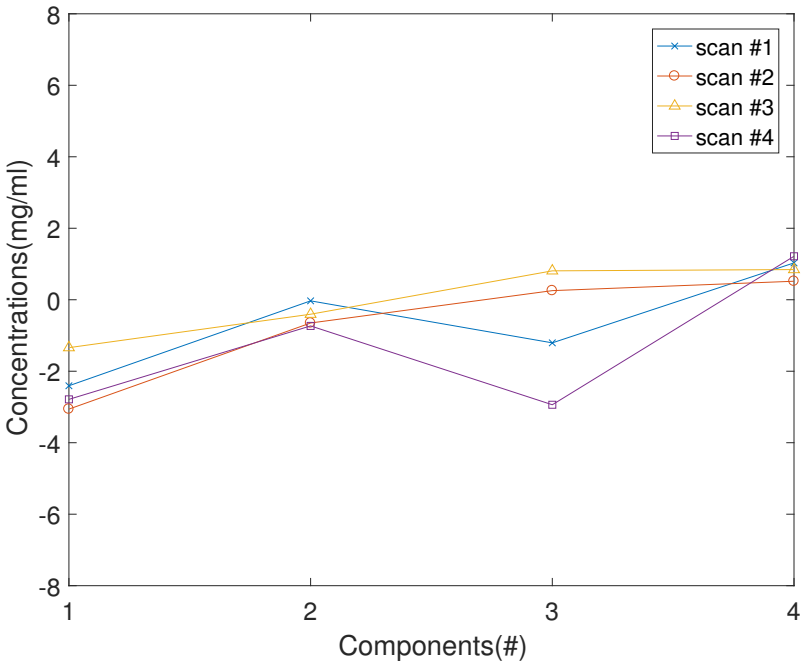


(c) Weighted concentration in SV from scan# 1

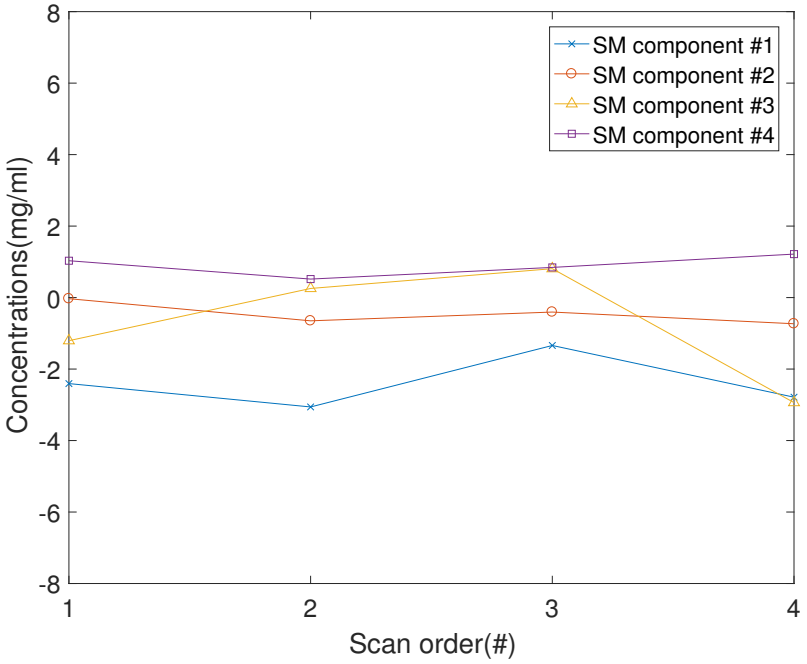


(d) Weighted concentration in SL from scan# 1

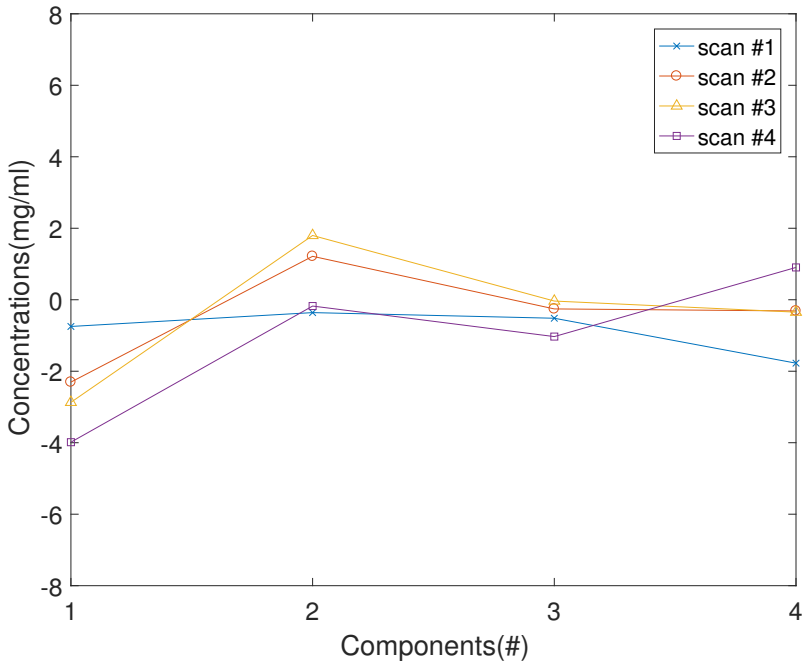
Figure 4.6: 3D rendering of weighted concentration(mg/ml) from scan #1



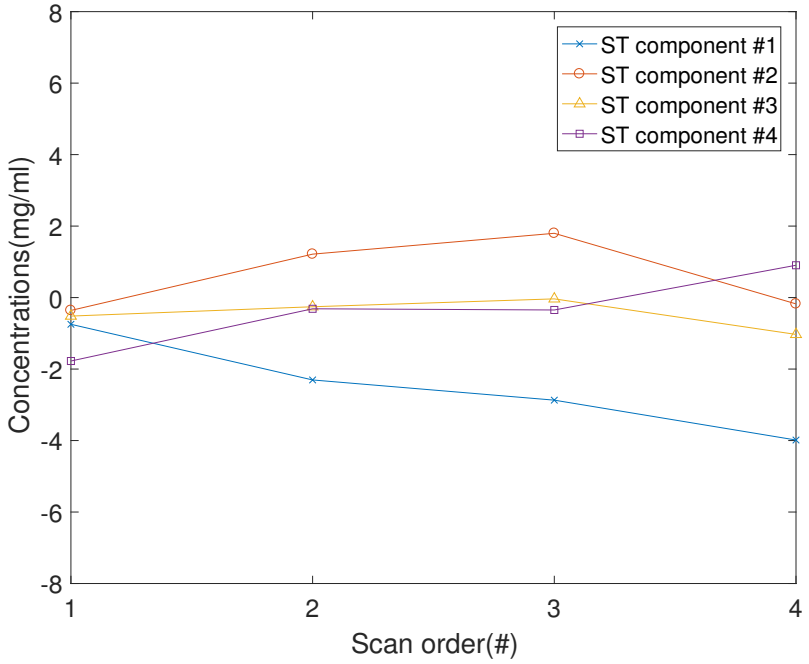
(a) The spatial distribution of weighted average concentrations in SM from scan #1-4 subtracted by concentrations of scan #0



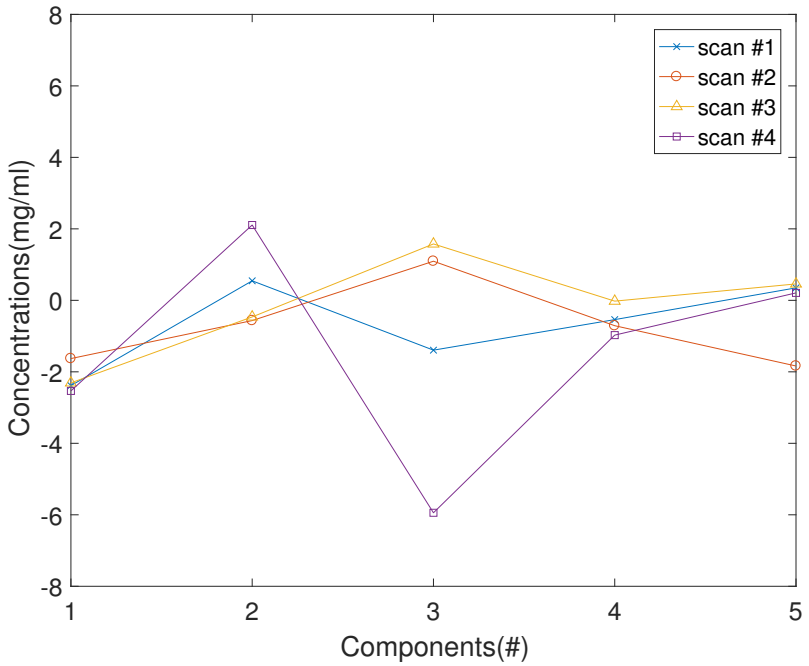
(b) The temporal distribution of weighted average concentrations in SM from scan #1-4 subtracted by concentrations of scan #0



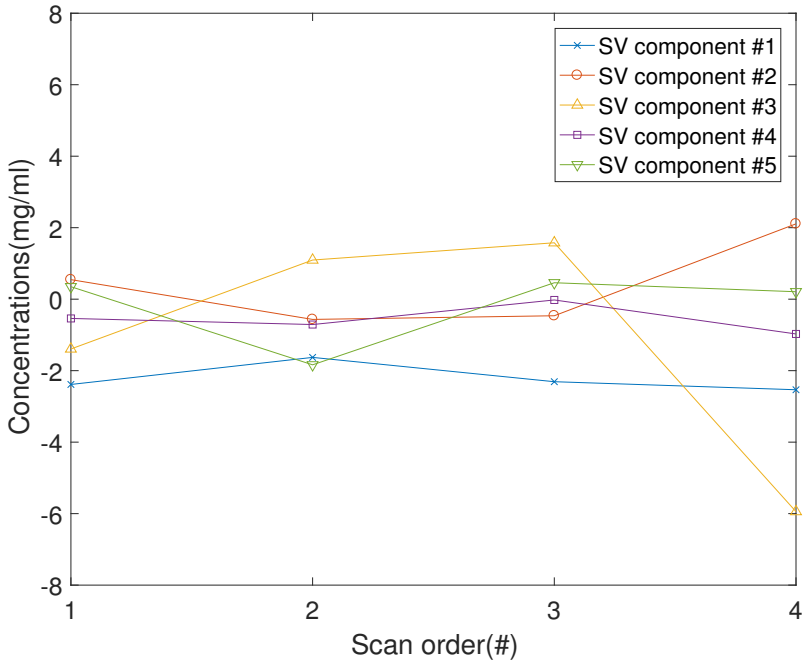
(c) The spatial distribution of weighted average concentrations in ST from scan #1-4 subtracted by concentrations of scan #0



(d) The temporal distribution of weighted average concentrations in ST fromscan #1-4 subtracted by concentrations of scan #0



(e) The spatial distribution of weighted average concentrations in SV from scan #1-4 subtracted by concentrations of scan #0



(f) The temporal distribution of weighted average concentrations in SV from scan #1-4 subtracted by concentrations of scan #0

Figure 4.7: The spatial and temporal distributions of weighted average concentrations in scalae compartments from scan #1-4 subtracted by concentrations of scan #0

4.4 Results and Analysis

The registration results of our approach appears to be good qualitatively and are quantified by the NCC metric value. Even though it is hard to determine in an absolute sense whether 0.467 is a good NCC value, the increase in NCC gives us a relative sense of the improvement between registration stages. To analyze the registration accuracy in a more meaningful quantitative way, ground truth needs to be established by labeling key points or 3D region, which is usually done manually.

Some features of the extracted concentrations distributions can be explained in terms of registration. In the concentrations distribution of scan #1 shown in Fig. 4.5, more abnormal high concentration vertices concentrated around the boundary of structures, which is actually caused by the registration error, and the high values come from the bone structures of cochlea that has much higher HU values. These errors also cause the apex component of each structure have higher average concentration because it has fewer vertices and thus is more sensitive to registration error.

From the baseline-benchmarked temporal concentration distribution shown in Fig. 4.7b, 4.7d and 4.7f, especially watching the component with highest number in each structure which is the most closed to where ioversol should enter, we can see that the change of concentrations are quite small compared with the concentration of injected ioversol and is mostly shown as fluctuations due to registration error. Because of this, we think that no ioversol entered the cochlea region in this experiment.

Chapter 5

Conclusion

In this thesis, we proposed a systematic approach to measure th concentration of contrast agent in the mouse cochlea by analyzing μCT images. We developed a 3D subject-atlas registration algorithm in ITK with multi-resolution and multi-stage strategies to segment and identify intra-cochlea structures from μCT images. The weighted average of concentrations based on weights computed by solving a Poisson equation and the volume discretization help to quantify the drug pharmacokinetics in scalae tympani (ST), scala vestibuli (SV) and scala media (SM).

An experiment on a series of five μCT scans shows the effectiveness of our system. A problem due to small movements of the mouse between scans is solved by an additional scan-scan registration process. An accurate 3D segmentation result is achieved and thus a full 3D concentration distribution in scalae compartments is obtained. We further quantify the concentrations in discretized volumes of each compartment for more specific estimates of concentrations.

Possible future research based on the proposed approach includes building a shape model using the segmentation result of cochlea structures, and using the segmentation result of our approach as the initial position of other advanced segmentation algorithms to achieve more accurate segmentation results.

References

[1] Arnold, W., P. Senn, M. Hennig, C. Michaelis, K. Deingruber, R. Scheler, H.-J. Steinhoff, F. Riphagen, and K. Lamm (2004). Novel slow-and fast-type drug release round-window microimplants for local drug application to the cochlea: an experimental study in guinea pigs. *Audiology and Neurotology 10*(1), 53–63. 1

[2] Bach Cuadra, M., V. Duay, and J.-P. Thiran (2015). *Handbook of Biomedical Imaging: Methodologies and Clinical Research*. Boston, MA: Springer US. 4

[3] Back, T. (1996). *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press. 14, 15

[4] Bertsekas, D. P. (1999). *Nonlinear programming* (2nd ed.). Athena Scientifi. 11

[5] Borkholder, D. A., X. Zhu, and R. D. Frisina (2014). Round window membrane intracochlear drug delivery enhanced by induced advection. *Journal of Controlled Release 174*, 171–176. 27

[6] Borkholder, D. A., X. Zhu, B. T. Hyatt, A. S. Archilla, W. J. Livingston, and R. D. Frisina (2010). Murine intracochlear drug delivery: reducing concentration gradients within the cochlea. *Hearing research 268*(1), 2–11. 1

[7] Brandt, A. (1977). Multi-level adaptive solutions to boundary-value problems. *Mathematics of computation 31*(138), 333–390. 39

[8] Braun, K., F. Böhnke, and T. Stark (2012). Three-dimensional representation of the human cochlea using micro-computed tomography data: presenting an anatomical model for further numerical calculations. *Acta oto-laryngologica 132*(6), 603–613. 2

[9] Dai, Y.-H. and Y. Yuan (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization 10*(1), 177–182. 13

[10] Dennis, Jr, J. E. and J. J. Moré (1977). Quasi-newton methods, motivation and theory. *SIAM review 19*(1), 46–89. 12

[11] Fletcher, R. and C. M. Reeves (1964). Function minimization by conjugate gradients. *The computer journal 7*(2), 149–154. 13

[12] Gorelick, L. Basic 3d geometric multigrid in matlab. 39

[13] Gorelick, L., M. Galun, E. Sharon, R. Basri, and A. Brandt (2006). Shape representation and classification using the poisson equation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*(12), 1991–2005. 22, 23

[14] Haghpanahi, M., M. B. Gladstone, X. Zhu, R. D. Frisina, and D. A. Borkholder (2013). Noninvasive technique for monitoring drug transport through the murine cochlea using micro-computed tomography. *Annals of Biomedical Engineering 41*(10), 2130–2142. 2, 3, 27

[15] Hartley, R. I. and A. Zisserman (2004). *Multiple View Geometry in Computer Vision* (Second ed.). Cambridge University Press, ISBN: 0521540518. 16, 20

[16] Hestenes, M. R. and E. Stiefel (1952). *Methods of conjugate gradients for solving linear systems*, Volume 49. NBS. 13

[17] Johnson, H. J., M. M. McCormick, and L. Ibanez (2015). *The ITK Software Guide: Design and Functionality* (Fourth ed.). ISBN 9781-930934-28-3. 5, 11, 36, 43

[18] Kanzaki, S., M. Fujioka, A. Yasuda, S. Shibata, M. Nakamura, H. J. Okano, K. Ogawa, and H. Okano (2012). Novel in vivo imaging analysis of an inner ear drug delivery system

in mice: comparison of inner ear drug concentrations over time after transtympanic and systemic injections. *PloS one 7*(12), e48480. 2

[19] Kanzaki, S., K. Watanabe, M. Fujioka, S. Shibata, M. Nakamura, H. J. Okano, H. Okano, and K. Ogawa (2015). Novel in vivo imaging analysis of an inner ear drug delivery system: Drug availability in inner ear following different dose of systemic drug injections. *Hearing research 330*, 142–146. 2

[20] Klein, S., M. Staring, and J. P. Pluim (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing 16*(12), 2879–2890. 14

[21] Mattes, D., D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank (2003). Pet-ct image registration in the chest using free-form deformations. *IEEE transactions on medical imaging 22*(1), 120–128. 20

[22] Mukai, H. (1979). Readily implementable conjugate gradient methods. *Mathematical Programming 17*(1), 298–319. 13

[23] Müller, M., R. Klinke, W. Arnold, and E. Oestreicher (2003). Auditory nerve fibre responses to salicylate revisited. *Hearing research 183*(1), 37–43. 1

[24] Noble, J. H., R. F. Labadie, O. Majdani, and B. M. Dawant (2011). Automatic segmentation of intracochlear anatomy in conventional CT. *IEEE Transactions on Biomedical Engineering 58*(9), 2625–2632. 2

[25] Nocedal, J. and S. Wright (2006). *Numerical optimization*. New York: Springer. 11, 12

[26] Plontke, S. K. and A. N. Salt (2003). Quantitative interpretation of corticosteroid pharmacokinetics in inner fluids using computer simulations. *Hearing research 182*(1), 34–42. 1

[27] Plontke, S. K., N. Siedow, R. Wegener, H.-P. Zenner, and A. N. Salt (2007). Cochlear pharmacokinetics with local inner ear drug delivery using a three-dimensional finite-element computer model. *Audiology & neuro-otology 12*(1), 37–48. 1

[28] Plontke, S. K., A. W. Wood, and A. N. Salt (2002). Analysis of gentamicin kinetics in fluids of the inner ear with round window administration. *Otology & neurotology 23*(6), 967–974. 1

[29] Pluim, J. P., J. A. Maintz, and M. A. Viergever (2003). Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging 22*(8), 986–1004. 9

[30] Polak, E. and G. Ribiere (1969). Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge 3*(1), 35–43. 13

[31] Poznyakovskiy, A. A., T. Zahnert, Y. Kalaidzidis, R. Schmidt, B. Fischer, J. Baumgart, and Y. M. Yarin (2008). The creation of geometric three-dimensional models of the inner ear based on micro computer tomography data. *Hearing research 243*(1), 95–104. 2

[32] Robbins, H. and S. Monro (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400–407. 13

[33] Rolston, L. and N. Cahill (2016). Interior and exterior shape representations using the helmholtz equation. *Proc. CompImage 2016 (Computational Modeling of Objects Presented in Images: Fundamentals, Methods, and Applications)*. 23

[34] Rueckert, D., L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging 18*(8), 712–721. 20

[35] Salt, A. N. (2002). Simulation of methods for drug delivery to the cochlear fluids. *Advances in Oto-Rhino-Laryngology 59*, 140–148. 1

[36] Salt, A. N. and Y. Ma (2001). Quantification of solute entry into cochlear perilymph through the round window membrane. *Hearing research 154*(1), 88–97. 1

[37] Santi, P. A., I. Rapson, and A. Voie (2008). Development of the mouse cochlea database (MCD). *Hearing research 243*(1), 11–17. 32

[38] Schwefel, H.-P. P. (1993). *Evolution and optimum seeking: the sixth generation*. John Wiley & Sons, Inc. 15

[39] Sederberg, T. W. and S. R. Parry (1986). Free-form deformation of solid geometric models. *ACM SIGGRAPH computer graphics 20*(4), 151–160. 20

[40] Sotiras, A., C. Davatzikos, and N. Paragios (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging 32*(7), 1153–1190. 20

[41] Studholme, C., D. L. Hill, and D. J. Hawkes (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition 32*(1), 71–86. 9

[42] Xianfen, D., C. Siping, L. Changhong, and W. Yuanmei (2006). 3D semi-automatic segmentation of the cochlea and inner ear. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 6285–6288. IEEE. 2

[43] Yoo, S. K., G. Wang, J. T. Rubinstein, M. W. Skinner, and M. W. Vannier (2000). Three-dimensional modeling and visualization of the cochlea on the internet. *IEEE Transactions on Information Technology in biomedicine 4*(2), 144–151. 2

[44] Yoo, S. K., G. Wang, J. T. Rubinstein, and M. W. Vannier (2001). Semiautomatic segmentation of the cochlea using real-time volume rendering and regional adaptive snake modeling. *J. Digital Imaging 14*(4), 173–181. 2

[45] Zou, J., M. Hannula, S. Misra, H. Feng, R. H. Labrador, A. S. Aula, J. Hyttinen, and I. Pyykkö (2015). Micro ct visualization of silver nanoparticles in the middle and inner ear of rat and transportation pathway after transtympanic injection. *Journal of nanobiotechnology 13*(1), 1. 2